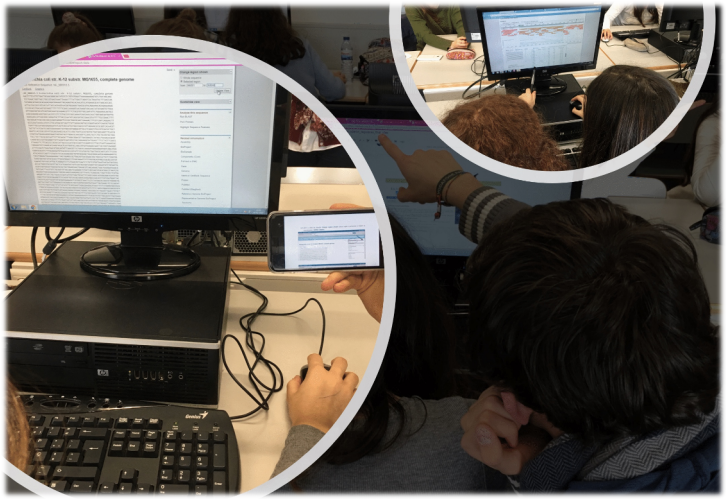


Integrating Bioinformatics at the Interdisciplinary Intersection of Elementary and Secondary Curricula using a Bottom-up Approach

Ana Sofia Ferreira Martins
Doutoramento em Biologia
Departamento de Biologia
2021

Orientador
Fernando Manuel dos Santos Tavares, Professor Associado
Faculdade de Ciências da Universidade do Porto (FCUP)

Coorientador
Leonor Mendes de Freitas Queirós e Lencastre, Professora Auxiliar
Faculdade de Psicologia e Ciências da Educação da Universidade do Porto (FPCEUP)

Integrating bioinformatics at the interdisciplinary intersection of elementary and secondary curricula using a bottom-up approach

The work presented in this thesis was carried out at *Faculdade de Ciências da Universidade do Porto* (FCUP) and *Microbial Diversity and Evolution Group* (MDE) of Centro de Investigação em Biodiversidade e Recursos Genéticos, InBio Laboratório Associado (CIBIO-InBIO), and financially supported by the Fundação para a Ciência e a Tecnologia (FCT) fellowship SFRH/BD/112038/2015.

Acknowledgments

During these four years I had the pleasure to have and keep by my side people who fully supported me and believe in me. This was what made this challenge possible in my life. Now, wearing a big smile, I feel that I am not only a better professional but also a better person.

Em primeiro lugar, o meu muito obrigada ao Exmo. Professor Fernando Tavares, por tudo! Pelas oportunidades que me proporcionou, que me fizeram crescer e ver novas perspetivas. Por ter aceite construir este trabalho comigo, por me ter inspirado e orientado de forma tão construtiva. O meu muito obrigada também por todos os momentos de reflexão, de conversa, de convívio e de sorrisos. Levo comigo ensinamentos para a vida. Ainda agradecer ao Professor e família que sempre tão bem me receberam. Com todos muito aprendi. Guardo-vos no coração.

À Exma. Professora Leonor Lencastre, por ter aceite este desafio, por toda a orientação e por sempre manter as portas abertas para mim. Por me colocar em contacto com oportunidades que em muito melhoraram a minha formação profissional e por, até ao fim deste projeto, ter comigo sempre partilhado pensamento positivos.

À Doutora Maria João Fonseca, também o meu muito obrigada. Pela inspiração e ajuda! O teu trabalho foi e é para mim um modelo a seguir. Tens uma simpatia imensa e uma disponibilidade e cuidado que admiro e que em muito contribuiu para este trabalho. O meu muito obrigada por teres sempre estado lá, por apoiares, acreditares e a qualquer hora em qualquer dia me teres ajudado!

À Exma. Professora Marina Lemos pela colaboração que em muito enriqueceu este projeto, o meu muito obrigada. Ainda, ao Exmo. Professor Patrício Costa por todo o apoio e pela oportunidade de formação.

Ao Exmo. Professor Luís Calafate gostaria de agradecer os comentários, as reflexões e os momentos de discussão acerca da formação dos “nossos” futuros professores.

À Exma. Professora Susana Pereira, por me ter apresentado ao Professor Fernando, e por me ter guiado até ao início deste caminho que agora se completa.

À Exma. Professora Conceição Santos, por me ter desafiado e por termos estabelecido tão importante colaboração nesta reta final. Estendo os meus agradecimentos à Célia, à Isabel e à Lina. Obrigada pela força!

Uma palavra de enorme agradecimento a todos os Professores que colaboraram neste projeto e que tornaram possível toda esta investigação: Professor Alexandre Ferreira,

Professora Ana Luísa Ferreira, Professora Ana Mesquita, Professora Cândida Ramoa, Professora Carmen Madureira, Professora Dalila Amandi, Professor Edgar Fernandes, Professora Fátima Marques, Professora Gabriela Chaves, Professora Graça Chaves, Professor Luís Moreira, Professora Luísa Santos, Professora Lurdes Cardoso, Professora Maria José Miranda, Professor Pedro Pimenta, Professora Sandra Ferraz; a todas as escolas envolvidas: Colégio Casa-Mãe, Colégio Salesianos do Estoril, Escola Secundária Alexandra Herculano, Escola Secundária da Boa Nova, Escola Secundária da Maia, Escola Secundária de Águas Santas; e a todos alunos que participaram neste projeto.

I also would like to say a big thank to Professor Jan Alexis Nielsen, Marc Rodemer, Susanne Jansen, Kirsi Ikonen, Rie Malm, Professor Andreas Haraldsrud and Johanna Taglieber for their availability to assist me in the analysis of the science curricula in their countries and for keeping me updated regarding the status of the integration of bioinformatics.

I am also grateful to all ELLS Learning Lab 2017 team, to EERA Summer School 2017 team, to ESERA Summer School 2018 team (particularly to Professor Gillian Roehrig and Professor Mathias Ropohl) and to Swiss Doctor School 2019 team. All of these events were unique experiences to improve my background, to open my perspectives, to stimulate my ideas and to grow.

Gostava ainda de agradecer a todos os alunos e monitores da Universidade Júnior (Pedro, Olena, Sofia, Sara e Rita) por terem colaborado e contribuído para este projeto.

E agora... à minha querida Leonor pelos momentos de descontração, pelo counselling, pela amizade, pelas divagações e pelos sorrisos e surpresas... por estes (quase) 4 anos de cumplicidade...o meu obrigada!

O meu obrigada aos elementos do MDE (atuais e aos que para mim serão sempre MDE): ao Doutor Pedro, à Cristina, à Rita, à Doutora Camila, à Mafalda, ao Nuno, ao Rafael, à Doutora Cláudia Serra, à Rafaela, à Sofia Lavrador, à Doutora Rose. Sem vocês no meu dia-a-dia, as coisas teriam sido bem mais difíceis! Obrigada por me receberem, por me fazerem crescer e por eu ter recebido alguns de vocês que muita dinâmica trouxeram ao grupo e ao meu trabalho. À Helena também o meu obrigada por desde sempre estares lá.

Dedico o último parágrafo destes agradecimentos aos meus alicerces. Ao Daniel, por ter feito parte desta fase desde o início, obrigada pela compreensão, pela calma, pelo encorajar, obrigada por estares aqui.

E aos meus pais…difícil resumir por palavras tudo o que sinto neste momento. São as minhas andorinhas que nunca deixarão a minha primavera acabar! Obrigada por sempre acreditarem em mim, por sempre estarem lá, por me abrirem portas, por me darem força e fazerem olhar para o futuro com os pés na Terra mas com a mente sonhadora. Obrigada é pequeno para tudo o que me dão.

À Minha Mãe e ao Meu Pai.

Ao Daniel.

Ao Sirius também.

"Aqueles que passam por nós não vão sós,
não nos deixam sós.
Deixam um pouco de si,
levam um pouco de nós..."

Antoine de Saint-Exupéry

Abstract

The new paradigms of biological research, in which bioinformatics plays a key role, cannot be disregarded when designing educational interventions. Therefore, they must be integrated into elementary and secondary school level as a didactic resource aimed to develop students' scientific literacy, through initiatives validated by science education research.

The knowledge regarding the impact of the integration of bioinformatics tools at the pre-university level, especially in Portugal, is limited. In addition, the characterization of the potential of bioinformatics tools as didactic resources according to teachers' perspectives is insufficiently described, namely in the national context. On the other hand, it is also essential to update the core curricular contents related to genomics and bioinformatics, particularly attending the need to overcome students' misconceptions.

Having in mind the needs and gaps identified in bioinformatics education, this study aimed to meet the urgent need to integrate bioinformatics at the elementary and secondary education level, updating both pedagogical practices and didactic resources, in which teachers and students have a central role.

The research started with the conception of a portfolio with selected bioinformatics activities, curricular-oriented, adjusted to class timetables, and focused on overcoming diagnosed misconceptions and in the introduction of new core concepts.

In order to assess the potential of bioinformatics as a learning tool, namely its impact on scientific and digital literacy, on the interest and attitudes of students, the activities were implemented both in the formal and non-formal educational context. The data gathered allowed to observe that students were curious about the bioinformatics and that, after a theoretical introduction about the potential and purposes of the bioinformatics web resources, they feel comfortable to explore them. Furthermore, the students showed to recognize the importance of bioinformatics for scientific research, its implications to society and to understand that most bioinformatics platforms do not require programming competencies, on the contrary, they are user-friendly and very intuitive for beginners. In addition, pupils reported they have a more concrete notion of research practices based on bioinformatics, ultimately contributing to motivate students for scientific careers. Students showed to be able to overcome misconceptions related to genomics and demonstrated a positive attitude towards the integration of bioinformatics in their learning practices, reinforcing its value in the educational context and as a promoter of citizenship education. These results reinforce the added-value of bioinformatics as a didactic tool that contributes to improve students' knowledge, skills

and motivation, which may enhance students' ability to overcome misconceptions, and to learn new core concepts.

The implementation of the activities also led to the definition of an inventory of fundamental and updated genomics-related concepts. This inventory is a resource that sheds some light on the need to update the science curricula, meeting the international recommendations for revising science school frameworks.

Depicted the impact of the designed bioinformatics activities on students' learning and interest, it became essential to characterize the teachers' perspectives about the benefits of integrating bioinformatics in elementary and secondary education, as well as about the main constraints that are preventing teachers from successfully implementing bioinformatics in their classrooms. The results revealed that the teachers were motivated and interested in learning more about bioinformatics and about strategies to integrate it in their pedagogical practices, recognizing that bioinformatics assumes a key role in scientific research and, therefore should be introduced in the science curricula. Teachers highlighted the adequacy of the exercises proposed to the curricular topics for elementary and secondary school level, as well as mentioned to feel more confident to approach bioinformatics in the classroom after attending bioinformatics training courses. The main constraints identified by the teachers in the implementation of bioinformatics-based activities were logistics conditions, and the language barrier since platforms are usually written in English. Although teachers recognized that their schools are equipped with computers and internet connection, they admit that computers are often obsolete, with outdated software and inadequate internet connections, impairing their use for bioinformatics-based activities. Additionally, teachers mentioned that the offer of bioinformatics courses for teachers is still scarce and asked for more training in this area. It was concluded that more training and long-term courses are needed to provide teachers with the knowledge and skills required to successfully and confidently introduce bioinformatics in their classes.

During the development of this project, a webpage was designed to be a repository of resources and an online forum for teachers to exchange experiences and doubts. The development of this website has gained particular relevance after teachers claimed for further support to implement bioinformatics resources in the classroom.

Taking into account the activities developed, the results obtained, and deliverables produced, this study is a contribution to underline the importance of integrating bioinformatics-based practices at the elementary and secondary school level. Furthermore, the research is also a strong assemble for future studies namely by

strengthening the portfolio of bioinformatics activities; extending the analysis to other students' populations; empower teacher training programs; and foster dissemination through the communication channels created during this project.

Keywords: Attitudes, Bioinformatics, Core Concepts, Elementary and Secondary School Level, Interest, Misconceptions, Scientific Literacy, Teacher Professional Development

Resumo

Os novos paradigmas de investigação biológica, entre os quais a bioinformática assume um papel fundamental, não devem ser ignorados no planeamento de intervenções educativas. Assim, estes devem ser integrados no 3º ciclo do ensino básico e no ensino secundário como recursos didáticos focados em desenvolver a literacia científica dos alunos, através de iniciativas validadas por investigações nas ciências da educação.

O conhecimento relativo ao impacto da integração de ferramentas bioinformáticas no ensino pré-universitário, especialmente em Portugal, é limitado. Para além disso, e na perspetiva dos professores, a caracterização do potencial das ferramentas bioinformáticas como recursos didáticos não se encontra suficientemente descrita, nomeadamente no contexto nacional. Por outro lado, é também fundamental a atualização dos conteúdos curriculares essenciais, em particular atendendo à necessidade de resolver conceções alternativas.

Tendo em consideração as necessidades e lacunas identificadas na educação em bioinformática, este estudo pretendeu responder à urgente necessidade de integrar a bioinformática no 3º ciclo do ensino básico e no ensino secundário, atualizando simultaneamente as práticas pedagógicas e recursos didáticos, em que os professores e os alunos são agentes fundamentais.

A investigação iniciou-se com a conceção de um portefólio com uma seleção de atividades de bioinformática orientadas de acordo com os projetos curriculares, adaptadas aos tempos letivos regulamentados, e focadas na resolução de conceções alternativas específicas e na introdução de novos conceitos essenciais a reter.

Com o objetivo de avaliar o potencial da bioinformática como ferramenta de aprendizagem, nomeadamente o seu impacto na literacia científica e digital, no interesse e nas atitudes dos alunos, as atividades foram implementadas em contexto formal e em contexto não formal. Os dados recolhidos permitiram observar que os alunos se revelaram curiosos acerca da temática “bioinformática” e que, depois de uma introdução teórica acerca do potencial e finalidades dos recursos de bioinformática, se sentem confortáveis em explorá-los. Mais ainda, os alunos mostraram reconhecer a importância da bioinformática para a investigação científica, a sua implicação no dia-a-dia da sociedade e compreender que a maioria das plataformas bioinformáticas não requerem competências de programação, pelo contrário, estas são *user-friendly* e muito intuitivas para iniciantes. Para além disso, depois das atividades, referiram ter uma noção mais concreta das práticas de investigação com base em bioinformática,

contribuindo em última instância para motivar os alunos para carreiras científicas. Os alunos mostraram ser capazes de resolver conceções alternativas relacionadas com a genómica e demonstraram uma atitude positiva relativamente à integração da bioinformática nas suas práticas de aprendizagem, reforçando o seu valor no contexto educativo e como promotor da educação em ciência para a sociedade. Estes resultados, reforçam o valor adicional da bioinformática como uma ferramenta didática que contribui para melhorar o conhecimento, competências e motivação, que podem potenciar a capacidade de os alunos resolverem conceções alternativas, e aprenderem novos conceitos essenciais.

Da implementação das atividades resultou ainda um inventário de conceitos fundamentais e atualizados relacionados com a genómica com o intuito de criar um recurso que orientasse a atualização do currículo escolar para as ciências. Esse inventário é uma ferramenta que demonstra a necessidade de atualizar os currículos de ciências, indo ao encontro das recomendações internacionais para a sua atualização.

Tendo-se caracterizado o impacto das atividades de bioinformática na aprendizagem e interesse dos alunos, tornou-se fundamental caracterizar a perspetiva dos professores acerca dos benefícios da integração da bioinformática no 3º ciclo do ensino básico e no ensino secundário, assim como acerca dos principais obstáculos à implementação da bioinformática na sala de aula. Os resultados revelaram que os professores estão motivados e interessados em aprender mais acerca da bioinformática e de estratégias para a integrar nas suas práticas, reconhecendo que esta assume um papel fundamental na investigação científica e, por isso, deve ser um ponto de partida para introduzir atividades curricularmente enquadradas. Os docentes enalteceram a adequação dos exercícios propostos aos tópicos curriculares para o 3º ciclo do ensino básico e ensino secundário, assim como mencionaram um aumento de confiança para abordar a bioinformática em sala de aula após a frequência de cursos de formação nesta área científica. Os principais constrangimentos identificados pelos professores na implementação destas atividades foram essencialmente de natureza logística, e as barreiras linguísticas, uma vez que as plataformas geralmente estão em inglês. Apesar de os professores reconhecerem que as suas escolas estão equipadas com computadores com ligação à internet, estes admitem que frequentemente os computadores são obsoletos, com *softwares* desatualizados e com ligações à internet inadequadas, comprometendo a sua utilização para atividades de bioinformática. Adicionalmente, estes professores mencionaram que a oferta de cursos de bioinformática para professores é escassa e, como tal, demonstraram o seu descontentamento, pedindo mais formação nesta área. Foi concluído que mais

formação e cursos de longa-duração são necessários para dotar os professores com o conhecimento e as competências necessárias para uma bem-sucedida e confiante introdução da bioinformática nas suas aulas.

No decurso deste projeto, foi ainda desenvolvida uma *webpage* com vista a criar um repositório de recursos e um fórum *online* para que os professores trocassem experiências e dúvidas. O desenvolvimento desta página web ganhou particular relevância atendendo à necessidade reportada pelos professores de se sentirem apoiados para implementarem recursos bioinformáticos na sala de aula.

Tendo em consideração as atividades desenvolvidas, os resultados obtidos, e produtos produzidos, este estudo é um contributo para o reforço da importância da integração da bioinformática no 3º ciclo do ensino básico e no ensino secundário. Mais ainda, a investigação é também um contributo para estudos futuros nesta área nomeadamente através do reforço do portefólio de atividades de bioinformática; pela extensão da análise a populações de estudantes mais diversas; assim como a continuação do investimento na formação de professores; e fomentar a divulgação através dos canais de comunicação criados ao longo do decurso deste projeto.

Palavras-chave: Atitudes, Bioinformática, Conceções Alternativas, Conceitos Essenciais, Desenvolvimento Profissional de Professores, Ensino básico e secundário, Interesse, Literacia Científica

Table of Contents

Chapter VI

List of Figures

List of Tables

List of Supplementary Files

List of Abbreviations

ACC – Acreditação de Ação de Formação

ANI - Average Nucleotide Identity

BLAST – Basic Local Alignment Search Tool

BSC – Barcelona Supercomputing Center

CCPFC – Conselho Científico-Pedagógico da Formação Contínua

EBI – European Bioinformatics Institute

EDGAR - Efficient Database framework for comparative Genome Analyses using BLAST score Ratios

ELLS – European Learning Laboratory for the Life Sciences

EMBL – European Molecular Biology Laboratory

FCT – Fundação para a Ciência e a Tecnologia

FCUP – Faculdade de Ciências da Universidade do Porto

GL4HS – GeneLab for High Schools

GMO – Genetically Modified Organism

HGP – Human Genome Project

HIV - Human Immunodeficiency Virus

IBM – International Business Machines Corporation

ICT – Information and Communication Technologies

IGC – Instituto Gulbenkian para a Ciência

JAX – The Jackson Laboratory

MaGe – Magnifying Genomes

MSc – Master of Science (Master's Degree)

NASA – National Aeronautics and Space Administration

NBIC – Netherlands Bioinformatics Centre

NCBI - National Center for Biotechnology Information

NGS – Next Generation Sequencing

NGSS – Next Generation Science Standards

NIH – National Institutes of Health

NLT – Nature, Life and Technology

OECD - Organization for Economic Co-operation and Development

ORF – Open Reading Frame

ORFfinder – Open Reading Frame finder

PhD – Doctor of Philosophy (Doctor's Degree)

PMP – Pathogen Modeling Program

SIB – Swiss Institute of Bioinformatics

SPSS – Statistical Package for the Social Sciences

STEM – Science, Technology, Engineering and Mathematics

TPACK – Technological Pedagogical Content Knowledge

TtGG – Teaching the Genome Generation

UJr – Porto's Junior University

UP – Universidade do Porto

Chapter I

General Introduction

General Introduction

1. Bioinformatics and its Importance for Scientific Research

In 1988, a special committee of the United States National Academy of Sciences started to think about the main aims of the Human Genome Project (HGP). Later on, in 1990, the initial research plan was defined and specific goals for the first five years were defined of what was then projected to be a 15-year research effort (National Human Genome Research Institute, 2019).

Eleven years after, in February 2001, the International Human Genome Sequencing Consortium published the first draft of the human genome (International Human Genome Sequencing Consortium, 2001; National Human Genome Research Institute, 2019). At that time, 90% of the sequence of the genome was completed. However the task ahead was to produce a finished sequence, by closing all gaps and resolving all ambiguities (International Human Genome Sequencing Consortium, 2001). The international team of researchers who aimed to sequence and map all the genes of *Homo sapiens* announced in April 2003 that a full sequence was published (Collins et al., 2003; National Human Genome Research Institute, 2019). Thus, once the sequence of all the DNA bases in the human genome was determined, maps that show the locations of genes for major sections of all chromosomes were produced, and linkage maps through which inherited traits can be tracked over generations were designed (National Human Genome Research Institute, 2019). It is also possible to read in the announcement of the publishing of the full genome of *Homo sapiens* that "*genomic era is* [now] *a reality*" and that "*a revolution in biological research has begun*". In fact, since the beginning of HGP, the projects' research strategies and experimental technologies generated a steady stream of ever larger and more complex genomic data sets that have been stored into public databases and have transformed the study of all life processes.

Since the beginning of the HGP until now, biological data are being produced at a phenomenal rate and computers have become indispensable to biological research. This fact is due to the improvements in computer technology which allows faster computations, better data storage and revolutionized the methods for accessing and exchanging data. Adding to this, the improvements in sequencing technologies turn sequencing cheaper and faster (Luscombe et al., 2001). The lower cost and quicker turnaround trend are continuing and the analytical capacity has become the limiting factor (National Research Council, 2012).

Currently, the higher quality, quantity and range of sequencing data confront scientists with big challenges (Lesk, 2019). In this context, bioinformatics is nowadays a key scientific tool in the genomic area. One of the aims of this scientific field is to organize data and submit new entries, to develop tools and resources that aid in the analysis of data and, to use tools to analyze data and interpret the results in order to depict their biologically meaning (Luscombe et al., 2001). Bioinformatics provides biologists with a markedly improved repertoire of research tools that allow them to study and understand the functioning of organisms at an unprecedented level of detail (Collins et al., 2003). Regarding bioinformatics definition, it is important to mention that it does not count on the agreement of the scientific community, especially in which concerns the difference between bioinformatics and computational biology. In 2002, Paulien Hogeweg of Utrecht University (the Netherlands) was invited by the Oxford University Press to prepare an entry with a definition of Bioinformatics taking into account that Oxford University Press considered her as the responsible for the term in her publication in cooperation with Ben Hesper in 1970 named: "*Bioinformatica: een werkconcept*" in which bioinformatics was defined as: "*the study of informatic processes in biotic systems"* (Hesper & Hogeweg, 1970; Hogeweg, 2011). Paired with this information, it is also possible to find in the literature the term "bioinformatics" coined to Deslisi and Bell in 1980 (Sadek, 2004). Trying to solve the challenge of defining bioinformatics and distinguish it from computation biology, in 2000 the U.S. National Institute of Health (NIH) published a working definition for bioinformatics and computational biology, recognizing that both disciplines are rooted in life sciences as well as computer and information science. According to NIH, Bioinformatics can be defined as: "*Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data including those to acquire, store, organize, archive, analyze or visualize such data*" (Huerta et al., 2000). Computational biology was defined as "*The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological behavioral and social systems*" (Huerta et al., 2000). Even with these contributions for a consensual definition of bioinformatics, and computational biology, since 2000 the scientific communities continue to discuss this topic which is mainly based on the fact that the two disciplines share the same roots and are partially and unavoidably interchangeable. In the context of this thesis, the NIH working definition of bioinformatics will be followed (Altman, 2009; Huerta et al., 2000; Murphy, n.d.).

Nowadays bioinformatics evolved in multidisciplinary areas such as functional genomics, proteome analysis, biodiversity, medicine, or drug design (Sadek, 2004).

Bioinformatics tools can be helpful, for example, to perform comparative genomic studies of bacteria like *Xanthomonas* strains to establish evolutionary relationships (Fernandes, 2019), or can be crucial in a completely different paradigm such as to optimize food preservation procedures or to avoid food fraud or to perform microbiome analysis (Dubey et al., 2020; Iannetti et al., 2017; Voorhuijzen-Harink et al., 2019). It is important to also highlight the high-value of the bioinformatics approach to medicine in which concerns cancer diagnosis and treatment within personalized and precision medicine (Adam, 2020; Coccia, 2020; Lafta, 2020; Patrinos, 2020). The efficient interpretation of *in silico* data both in biomedical research and in the healthcare systems opens new perspectives for a better understanding of diseases and development of better and personalized diagnostics and therapeutics, contributing to find answers and put forward therapeutic solutions for rare diseases (Belmont & Shaw, 2016; Merelli, Pérez-Sánchez, Gesing, & D'Agostino, 2014). Bioinformatics also contributed to optimize drug design especially in the area of Pharmacogenomics in which bioinformatics tools lead to a better understanding of the predicted effect of a specific drug in a target biomolecule, such as a protein, allowing the optimization of the *in vitro* and *in vivo* tests of drugs (Thorn et al., 2010).

Despite all the promising contributions of bioinformatics, is obvious that the future of this discipline is based on its combination with wet-lab research. In fact, new computational capabilities will enable the generation of hypotheses, and simulate the development of experimental approaches, to test them leading to experimental data which, in turn, will be used to generate more refined models that will improve overall understanding and increase opportunities for applications in different fields (Collins et al., 2003).

Another key point to support the potential of bioinformatics in future research is to give specifically dedicated training aiming to equip life sciences workforce to interpret and harness bioinformatics data and operate diverse types of bioinformatics software (Porter & Smith, 2019). Educational institutions need to address current knowledge gaps through interdisciplinary curricula, which should be started even during high school, complemented with dedicated training for teachers (Porter & Smith, 2019). Meeting these challenges will require skills to understand biological systems and to use information (Collins et al., 2003). In this regard, it is important to develop and implement training programs for scientists and to think about strategies to integrate bioinformatics in undergraduate curricula. It is reported that hands-on training on bioinformatics for biologists can have an important role to clarify the perceptions about bioinformatics tools

use and also to enable biologists to make informed decisions about how to use them (Anderson et al., 2007; Shachak et al., 2007; Shachak & Fine, 2008).

The changes in biological research based on bioinformatics procedures are designing future trends for this field, which has gone itself through a number of transformations during the past 20 years establishing itself as a key component of new biology (Ouzounis, 2012). The expansion of the field is creating prospects for novel employment opportunities and at the same time contributing to a variated range of scientific disciplines (Ouzounis, 2012; Porter & Smith, 2019).

The future challenges in this area are to create, enhance and implement new algorithms for data mining and to develop databases and web tools to access them (Mathur, 2018). Being aware of all the potentialities that bioinformatics offers, this scientific area will continuously be a daily basis practice in research in the next decades.

2. General Population Literacy on Bioinformatics

According to the key role of bioinformatics for scientific research and considering its implications for our daily life decisions as citizens, creating opportunities for society to improve their literacy on bioinformatics in the context of responsible innovation and sociotechnical knowledge through educational materials is of an immediate necessity (Brazas et al., 2014; Dressler et al., 2014; Porter &Smith, 2019; Whitley et al., 2020). Traditionally, bioinformatics tools and training programs have focused on life science audience (European Bioinformatics Institute, 2019; Swiss Institute of Bioinformatics, 2019; University of Cambridge, 2019). Consequently, bioinformatics is still a new topic for the general population. When discussing about general population, it is important to target who this "general population" is and, according to that, meditate on what "*should everyone know about bioinformatics?*" (Figure I - 1) (Cham, 2013). This thought was a motto for a communication of Hienke Sminia from the Netherlands Bioinformatics Centre in 2013 at the 21st International Conference on Intelligent Systems in Molecular Biology & 12th European Conference on Computational Biology (ISMB/ECCB 2013). She pointed out that the public audience is broad and can be divided into distinct age groups, each with unique interests and different backgrounds or awareness needs. In this scope, bioinformatics materials and activities developed for the public should be designed according to the target age (Brazas et al., 2014).

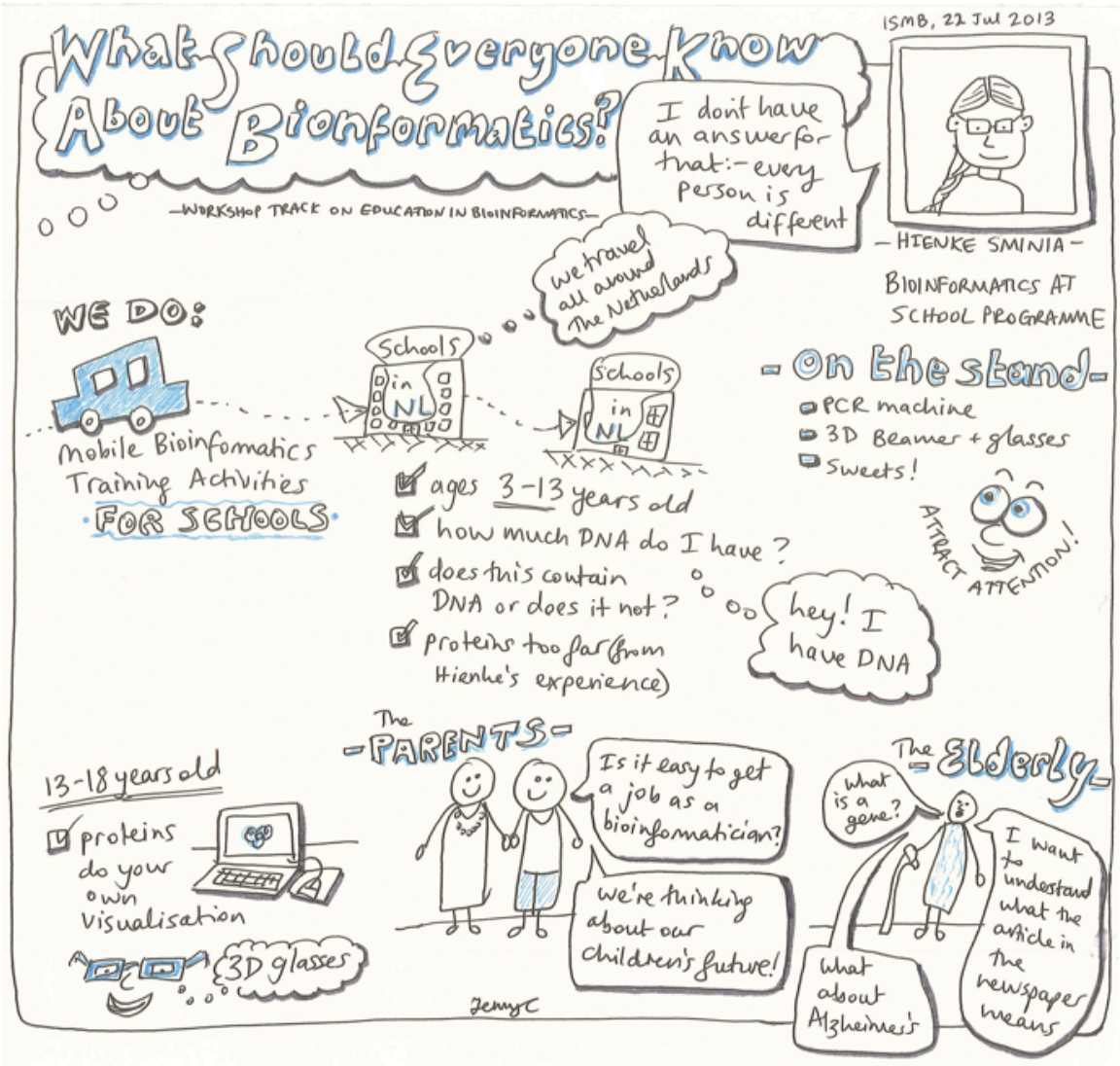


Figure I-1. A sketch of Hienke Sminia's presentation on the genomics and bioinformatics needs of the public audience from the ISMB/ECCB 2013 conference.

What is common to the general public interest is therefore the implications that bioinformatics and genomics can have in their daily life. An educated population would be available to understand the implications of these scientific areas in their healthcare, for example, comprehending the scientific foundations that support public policies. Having this in mind, a model was proposed by Machluf, Tal, et al. (2017) to be used by policymakers, caregivers, and health organizations to improve decision-making based on data. The authors claim that we are living in an evidence-based generation, in which data are privileged to support decisions. In the particular case in which data are related to health and health care, the decisions made based on these evidence have an impact on individuals and, consequently, on the general public (Machluf, Tal, et al., 2017). This example underlines the importance of data in public policies, reinforcing the need to get society acquainted with data and data analysis, namely by increasing individuals' literacy in bioinformatics.

Bioinformatics education is primordial in the biological research paradigm we are living in. More than being aware of the lack of population' literacy on genetics and genetic risk, society is rushed into misinformation regarding serious ethical issues through media sources. Recent examples of public news that were topics of discussion to the general public included for example when, in 2008, Time considered as the best invention of the year the kit commercialized by the company 23andMe which promised to the buyer to know "*the digital manifestation of you*" by an affordable price and so the buyer can learn and share his/her genetic secrets having as a motto "*It's all this information beyond what you can see in the mirror*" (23andMe, 2020; Hamilton, 2008). More recently (2020), this company was noticed because it sold the rights to another company to develop a new drug sharing the genomic database of their costumers, losing the confidentially of this information (Boyd, 2020; Hamzelou, 2020). Another example of sharing of personal genomic data was the program of National Geographic "*The Genographic Project*®" which aimed to reveal patterns of human migration (National Geographic, 2016). In both examples, people were invited to give their consent regarding their personal genomic data in order to know more about their genetic risk factors and their ancestrality. However, these initiatives did not provide training on strategies for data mining.

Having in mind that these initiatives are open to everyone, it is important for society and for policymakers to be aware that people cannot decide on topics they are not knowledgeable about. Therefore, there is an immediate need for the general public to improve their literacy on bioinformatics, which represents new and exciting challenges for bioinformatics training programs (Brazas et al., 2014; Whitley et al., 2020).

Globally, institutions believe that the first step to improve public scientific literacy on bioinformatics is to formally integrate bioinformatics in schools (Brazas et al., 2014). As claimed by Francis Collins et al. in 2003, we are in an "educable era" about genomics, and "*high-school students will be both the users of genomic information and the genomics researchers of the future*" (Collins et al., 2003; Whitley et al., 2020). In this regard, bioinformatics institutes and universities worldwide have now available bioinformatics-based programs to narrow the gap between the research and schools and also for a broader audience including the general population with time-limited experiences – non-formal education.

For example, the Netherlands Bioinformatics Centre (NBIC) has been promoting a program called BioWISE aimed to offer high-quality teaching in bioinformatics at all levels of education, promoting an initiative called Bioinformatics@school, in cooperation with other partners (Netherlands Bioinformatics Centre, 2019a, 2019b; Van Gelder et al., 2019). Another example is the program of the National Aeronautics and Space

Administration (NASA) Ames Research Center called GeneLab for High Schools (GL4HS). This is a four-week intensive training program for high school juniors and seniors which provides students with an opportunity to immerse themselves in space life sciences with a specific focus on omics-based bioinformatics research, the science of collecting and analyzing complex biological data such as genetic codes, and computational biology (National Aeronautics and Space Administration, 2019). A Portuguese example was the project "Bioinformatics at School (Bioinformática na Escola)" from Instituto Gulbenkian para a Ciência (IGC) aimed to promote bioinformatics to elementary and high school students (Fundação Calouste Gulbenkian - Instituto Gulbenkian de Ciência, 2012; Marques et al., 2014).

Adding to this and being aware of the importance of bioinformatics to medicine and to economy, universities in general and medical schools in particular should acknowledge the need to train their undergraduates in basic bioinformatics analysis and address ethical, legal, and social implications of genome research (Kesh & Raghupathi, 2004). As described by Kesh and Raghupathi (2004) the ownership of the human genome is probably the most critical issue.

Bioinformatics is growing in its interdisciplinarity and therefore it is crucial to train different audiences in this field (Gurwitz, Weizman, & Rehavi, 2003; Juran & Server, 2001; Parsons, 2002). Another interesting program is promoted by Barcelona Supercomputing Center's (BSC) and it is called "Bioinfo4Women: Outstanding Young Female Bioinformaticians" based on gender equality in science, promoting the contribution of women researchers in the bioinformatics area (Barcelona Supercomputing Center, 2019).

Strategies to integrate bioinformatics-based approaches in education settings, both formal and non-formal, should be designed and implemented as a helpful tool to improve general public knowledge on this scientific topic and at the same time, motivating students to proceed a Science, Technology, Engineering and Mathematics (STEM) career which is simultaneously based on biology and computational technology (Athanasiadis et al., 2016; Kovarik et al., 2013; Machluf & Yarden, 2013; Martins et al., 2020a; Whitley et al., 2020).

3. Bioinformatics Education: Curricular Framing

Aiming to increase public literacy in bioinformatics, a portfolio of studies and projects were and still are being performed in order to highlight the added value of integrating bioinformatics in both formal and non-formal educational contexts. As mentioned previously, integrate bioinformatics-based approaches in school will be an excellent

contributor to reach the general population and to contribute to a well-informed society that can make informed decisions when it comes to a bioinformatics-based issue (Athanasiadis et al., 2016; Brazas et al., 2014; Lewitter & Bourne, 2011; Whitley et al., 2020).

It is reported in the literature the urgency of updating the life science curricula in order to include bioinformatics-based educational approaches. Bioinformatics is especially well-suited to fulfill the requirements of an educational instrument since it may attend many aspects of biology curricula, it is transdisciplinary and it supplies a platform for inquiry-based learning in the classroom, empowering students to access and use real scientific data (Cummings & Temple, 2010; Wood & Gebhardt, 2013).

This curricular update should consider the scientific methodology and define core competencies on bioinformatics education (Sayres et al., 2018; Welch et al., 2014). Simultaneously, and inspired by daily scientific advances, this urge is not only reported to life science curricula but also for mathematics, for example (Boaler & Levitt, 2019; Wightman & Hark, 2012). In an opinion article at Los Angeles Times Journal it is discussed by Boaler and Levitt that curricula should be updated in order to teach about data science and not about traditional mathematical approaches, i.e., data should be put in discussion in the classroom and its analysis at the center of high school mathematics scaffolding students' comprehension of how applied math's may address real-world issues (Boaler & Levitt, 2019).

Being aware of this need, the scientific community already started to work with educational stakeholders in order to promote bioinformatics integration in the classrooms, and in the curricula. In Denmark, after the updates on the sciences curricula for high school of 2017, bioinformatics became a core content in the guidelines for Biology A and Biotechnology A of the academic high school branch (stx), for the topic: Genetic and Molecular Biology addressing gene technology and bioinformatics (Børne - og undervisningsministeriet, 2017a, 2017b).

In Norway, although bioinformatics has not been officially integrated into curricula, from 2017 until 2020, there was a pilot testing curriculum for a discipline called "*Programming and modeling X*", in which students were introduced to programming and modeling and it was possible to use bioinformatics to tackle biological issues. Students were given the opportunity to model population growth, simulate vaccination effects and work with large datasets (e.g. *in silico* translate DNA into RNA and proteins) (Haraldsrud, 2017; Husum & Haraldsrud, 2017).

In the Netherlands, despite bioinformatics has not yet been acknowledged as a topic in the biology curricula, there is consensus about its importance, so teaching modules of bioinformatics such as "*Bioinformatica: moleculaire biologie achter de computer*", designed by Wageningen University in 2014, have been created and are used in some schools (Wageningen University, 2014). This module was created framed in an initiative to promote "*Nature, Life and Technology* (NLT)". NLT is an integrated science and mathematics subject, which has been introduced for senior general secondary education (10/11 HAVO) and pre-university education (10/11/12 VWO). The general aims of NLT are to make science and mathematics education more attractive and challenging by offering students insights into new – often interdisciplinary - developments in science and technology and to create coherence in teaching and learning the science and mathematics subjects (Vereniging NLT, 2014).

The examples mentioned above are a model and an opportunity for the educational community to think about bioinformatics integration. Moreover, educators should also be aware that efforts can be made to promote bioinformatics-based approaches framed within the life sciences curricula that are being implemented. For example, in the United Kingdom, there is opportunity to integrate it in the guidelines of the national curriculum in England: science curriculum of study - key stage 4 (Department for Education (UK Government), 2014). According to the curriculum, students at this stage should be taught about "*the potential of genomics on medicine*" and about "*the uses of modern biotechnology including gene technology; some of the practical and ethical considerations of modern biotechnology*". To approach these topics, it is clear the potential of bioinformatics integration.

In Germany, there are 16 different federal states and they all have their own curricula. This particularity is a barrier to a standardization of the education policies and curricular orientations to be applied. Having an overview regarding the official guidelines that are common to all states and being aware that informatics is not a subject in every state, but biology is, it is possible to conclude that Genetics is the topic in which bioinformatics-based approaches could be integrated. In the document of the general requirements for biology exam in Germany, it is possible to list topics in which bioinformatics will be naturally integrated such as DNA and gene regulation, Polymerase Chain Reaction (PCR), genetic profile, among others (*Einheitliche Prüfungsanforderungen in Der Abiturprüfung Biologie*, 2004).

On the Finnish curricula, bioinformatics is also not explicitly mentioned in middle school or secondary school curricula in biology. However, studies aimed to identify the possibility of integrating bioinformatics in the upper secondary school curriculum already

started to be made, and indications were obtained claiming that: "*The role of bioinformatics in current upper secondary school biology education was found to be varying. The biology curricula contain multiple educational aims in which bioinformatics could be used as a beneficial aid. However, bioinformatics is not directly mentioned in the curricula*" (Pirinen, 2019).

In Austria, there are different school types for upper secondary school (from 14-18/19 years). Focusing on the curriculum for regular secondary school (Allgemeinbildende Höhere Schule (AHS)), chosen by the majority of the students, there is not any clear mention of bioinformatics (Bundesministerium für Bildung Wissenschaft und Forschung, 2019). Despite that, there are curricular requirements in which bioinformatics could be integrated. For example, one of the seven core concepts defined in the general aims of the curriculum is that students should obtain basic biological understanding, enabling students to evaluate biological findings/insights and consequently to engage/participate in the social discourse. It is also mentioned in the curriculum that biotechnological procedures should be discussed as well as their application and possible effects/consequences referring to ethics in science and bioethics. These are topics perfectly suitable to approach through bioinformatics-based exercises.

In Portugal, bioinformatics is not formally part of the life science curriculum for elementary or secondary education. However, as it happens in other countries as described before, there are key topics in which bioinformatics can be easily integrated namely genetic regulation or evolution (Martins et al. 2018, 2020a; Martins & Tavares, 2018; Mendes et al., 2003, 2004; Ministério da Educação, 2018a, 2018b).

In the United States, bioinformatics is also not explicitly in the Next Generation Science Standards (NGSS, 2013). However, it is possible to use bioinformatics platforms as tools to teach and to learn about topics of the curriculum such as to accomplish the Science and Engineering Practice of the topic "HS-LS3-3 Heredity: Inheritance and Variation of Traits" in which students should apply concepts of statistics and probability to scientific questions and problems, using digital tools (NGSS, 2013; Porter & Smith, 2019).

A study aimed to analyze the secondary school science standards of 49 U.S. states for bioinformatics related content found a generally low representation of bioinformatics-related content. Evolution is the highest represented category in opposite with Human Genome Project/genomics which was the lowest represented category among the 49 states. This research performed by Wefer and Sheppard (2008) draws recommendations

for reworking/rewording existing standards in order to promote science literacy in high schools (Wefer & Sheppard, 2008).

As claimed by Porter and Smith (2019), future versions of the Next Generation Science Standards (NGSS) should include bioinformatics so that most of the high schools would be using bioinformatics-based content. Students can be instructed to locate and analyze existing data on interesting biological phenomena produced by the scientific communities, opening opportunities to enhance their motivation and to make new discoveries ultimately endorsing citizen science practices (Chiovitti et al., 2019; Kawrykow et al., 2012; Kelling, 2012; Porter & Smith, 2019).

Overall, the official integration of bioinformatics in the curricula is at different stages in different countries. As reported before, it is possible to conclude that while bioinformatics is being parsimoniously integrated into the official curricula for middle and high school in some countries, which may foster similar initiatives in countries where efforts have to be made to integrate bioinformatics in teaching practices.

Being aware of this reality, there is a main goal that is common to all educational realities: the development of strategies that narrow the gaps between research and its impact on learning. Public and private educational institutes need to adapt to the changes in biological research, i.e., the share and real-time access of large amounts of biological information - big data, changed the experimental routines. Bioinformatics is in the spotlight to accomplish this aim, through formal and non-formal transdisciplinary strategies.

In this regard, it is not surprising that didactic resources are being designed, implemented and published in order to give teachers the tools to more easily integrate bioinformatics.

4. Bioinformatics Education: Didactic Resources

To support the changes needed to integrate bioinformatics in the curricula of middle and secondary school level, educational stakeholders should deliverable an action plan to meet teachers and students' needs (Attwood et al., 2017). This action plan should include an extensive survey of the available open access and user-friendly bioinformatics platforms and tools, in order to design and develop bioinformatics resources for schools aiming to empower bioinformatics education (Machluf & Yarden, 2013). Currently, teachers using bioinformatics resources in their classes is growing as well as the number of activities available (Machluf, Gelbart, et al., 2017), but they should be designed regarding learning, setting realistic and practical expectations. At the same time, learning environments should set specific and clear goals, based on real scenarios, and focused

on curricular contents being taught to conciliate the understanding of applied research and their learning achievements, being easily welcome by teachers (Form & Lewitter, 2011; Kovarik et al., 2013; Machluf, Gelbart, et al., 2017; Martins, Fonseca, et al., 2020a, 2018).

Moreover, when designing educational bioinformatics didactic resources, it is important to acknowledge the frequent misconceptions in order to explore the bioinformatics tools with the aim to overcome them. For instance, some misconceptions mainly related to genetics are a good starting point when trying to integrate bioinformatics. Recent findings revealed students' misconceptions related to gene definition or the role of genetic material, using indistinctly terms such as gene, genome, and genetic information without being certain about the scientific amplitude of their specific meaning (Bahar et al., 1999; Duncan et al., 2009; Lewis & Kattmann, 2004; Martins, Fonseca, et al., 2020a; Stern & Kampourakis, 2017). Moreover, some studies reported that students do not seem to clearly distinguish between phenotype and genotype, considering genes, traits, and characters as equivalents (Duncan et al., 2009; Lewis & Kattmann, 2004; Wood-Robinson et al., 2000).

Having these misconceptions in mind, it is also important to mention that, apart from the bioinformatics-related contents that are already mentioned in some curricula as explored in the previous section, new core concepts need to be introduced to fully understand bioinformatics-based activities. Some examples of these notions are "*Open Reading Frame*" (ORF) or "*Basic Local Alignment Tool*" (BLAST), that are genomics concepts well within the scope of secondary school biology education (Altschul et al., 1990; Martins & Tavares, 2018; nature.com, 2020; NCBI, 2020a, 2020b).

Being aware of the misconceptions and the new core concepts previously mentioned is fundamental to implement an assortment of bioinformatics-based activities. Several activities were published aimed to approach different areas of biology, from comparative genomics (Alaie et al., 2012; Bacusmo et al., 2019; Martins, Fonseca, et al., 2018), to pathogenic identification (Valenzuela et al., 2019), to study of protein structure and drug design (Eurich et al., 2012; Szarecka & Dobson, 2019; Taylor et al., 2014), to understand genome assembly (Taylor et al., 2013), genetic and gene identification (Martins, Fonseca, et al., 2018; Wefer, 2003; Yang et al., 2017), to DNA barcoding and ecology (Eble & Pecore, 2019; Rahn et al., 2019), to food safety studies (Fernandes et al., 2014; Martins, Lencastre, et al., 2018b) or to evolution (Bokor et al., 2014; David, 2018; Duffus, 2019).

Despite the vast portfolio of bioinformatics-based exercises available for classroom settings, the educational added-value of these activities is still poorly characterized regarding the dimensions of students' perceptions and attitudes (Wefer & Anderson, 2008). There is more to be understood regarding activities assessment. Some studies have reported that despite the assessment of bioinformatics activities revealed knowledge enhancement, affective changes, or skill acquisition, there are still gaps in the methodology of assessment that have to be revised in order to increase the robustness of the results obtained (Campbell & Nehm, 2013; Magana et al., 2014).

In fact, the assessment of bioinformatics-based activities on students concerning literacy, attitudes and interest, is crucial to successfully incorporate bioinformatics into high school science classrooms. Gelbart and Yarden (2006) conducted a research aimed to use bioinformatics-based tools for students to explore a genetics problem-solving approach in which BLAST tool is used to identify a mutated gene. They concluded that learning mediated by a bioinformatics environment promotes students' comprehension of scientific practices and reasoning (Gelbart & Yarden, 2006). Later on, in 2017, Machluf, Gelbart et al. (2017) underlined that students' cognitive outcomes revealed learning gains in bioinformatics, placing bioinformatics as part of students' scientific "toolbox". Furthermore, analysis of students' affective outcomes revealed positive attitudes towards bioinformatics and the learning environment (Machluf, Gelbart, et al., 2017). These outcomes have been corroborated by other research groups, namely by Marques et al. (2014) who conducted a pilot-study in Portugal with high school students which revealed that students find the activity enjoyable and observed a clear positive effect in both students' knowledge and confidence (Marques et al., 2014). Other studies emphasized that bioinformatics-based activities lead to significant gains in awareness, relevance, self-efficacy and engagement when trying to promote STEM careers (Kovarik et al., 2013; Whitley et al., 2020). Recently, Martins, Fonseca et al. (2020a) demonstrated that bioinformatics-based interventions improve students' knowledge of concepts, such as gene, start and stop codons, genetic code, and genomic or comparative genomics, as well as increase students' awareness about the applications and potential of bioinformatics.

Having in mind the panoply of bioinformatics-based activities available, covering a broad spectrum of scientific topics, learning environments and potential for middle and secondary education, the role played by teachers, the main agents to promote educational changes, remain poorly elucidated. It is well known that engaging directly with teachers shortens the time it takes to bring new scientific findings to the classroom (Wood & Gebhardt, 2013; Porter & Smith, 2019).

5. Teachers Role on Bioinformatics Education: Training and Needs

As detailed in the previous section, several bioinformatics-based didactic resources are currently available for teachers. Although, these are only hardly used by teachers in their classes, mainly because most teachers do not feel prepared to explore these resources without tutorial assistance and have no time to prepare didactic instruments attending their tight schedule (Machluf & Yarden, 2013; Martins et al., 2017, 2020b, 2020c; Martins, Lencastre, et al., 2018a; Moore, 2008). This reality is aligned with the fact that only a few in-service science teachers have received bioinformatics and genomics training or guidance on how to teach the subject to their students (LaRue et al., 2018). An additional constrain is that the majority of the resources available are in English, therefore not really accessible to teachers who are not English speakers (Machluf & Yarden, 2013; Martins et al., 2017; Martins, Lencastre, et al., 2018a; Martins et al., 2020b; Moore, 2008). Taken altogether, the evidence suggests that in order to integrate bioinformatics in the classrooms, it is required the involvement of policymakers and universities to design and promote strategies to refresh teacher knowledge (LaRue et al., 2018; Wray, 2017).

To scaffold teachers to use bioinformatics in their teaching practices different interventions are required. Firstly, teachers should be inquired regarding their perceptions about the relevance of bioinformatics for elementary and secondary education and also about their training needs in this scientific area. These assessments will certainly shed some light on the best way to enhance students' knowledge, interest, and attitudes regarding bioinformatics, and ultimately will contribute to identify key points of reflection for educational policymakers. Secondly, increase the offer of training courses for in-service teachers, focused on bioinformatics integration in the classroom. Without training and updated curricula, teachers feel uneasy to implement new teaching practices (Whitley et al., 2020). Based on previous assessments of teachers perceptions about bioinformatics integration in schools, it is clear that bioinformatics training courses, which combine knowledge, skills, and pedagogy in technology-rich environments, are infrequent and rare (Machluf & Yarden, 2013; Martins et al., 2017; Martins, Lencastre, et al., 2018a; Martins et al., 2020c). The lack of training opportunities is accompanied by the fact that many science teachers got their academic teaching degree at a time in which computers were rarely used or inaccessible for most people. It is therefore with no surprise that in-service teachers feel uncomfortable or not motivated to use computers

in their classrooms, being urgent to scaffold teachers to develop their Technological Pedagogical Content Knowledge (TPACK) (Koehler & Mishra, 2009; Mumtaz, 2000; Palha, 2019). Increasing, the TPACK of teachers in relation to bioinformatics, especially regarding the integration of domain-specific knowledge with procedural skills and the ability to envision its role in society, will help mobilize students' higher-order thinking skills (Wefer & Anderson, 2008). In fact, teachers can have a great impact on increasing public literacy in genomics and bioinformatics. Therefore, training programs should focus on exploring modern biology techniques, resources, and tools such as web-based bioinformatics activities to foster teachers' digital literacy and decreases their fear to use updated scientific resources such as databases and analysis tools (LaRue et al., 2018; Wood & Gebhardt, 2013). For instance, in Denmark, there are clear guidance for high school teachers to boost bioinformatics implementation in the Biology and in the Biotechnology curricula. Teachers are encouraged to approach bioinformatics by providing examples of bioinformatics databases, by comparing and analyzing nucleotide, protein sequences, mutation types and restriction sites and by thematizing evolutionary issues, for example (Undervisningsministeriet & Styrelsen for Undervisning og Kvalitet, 2018a, 2018b).

Recognizing the importance of the collaboration between research institutions and teachers in schools, the current offer of training courses for teachers is increasing, but it is still insufficient. Among the present offer of training courses for biology teachers, some examples are worth to be described.

"*Teaching the Genome Generation*" (TtGG) is a high school teacher professional development program promoted by The Jackson Laboratory (JAX) and funded by an NIH Science Education Partnership Award. This training program integrates instruction in molecular genetics laboratory techniques, bioinformatics and bioethics, mainly related to human genetic diversity and genomic medicine. An assessment made after four years of training course implementation revealed that TtGG increased teachers' abilities to integrate into their high school classrooms complex concepts about genomics. In the program of this training course, all the contents were discussed with teachers to optimize lesson plans and pedagogical strategies, which helped to design guidelines both for teachers and for students curricularly framed in order to facilitate the integration of the proposals in classrooms without being an extra burden for teachers (LaRue et al., 2018). When designing a training course for teachers it is important to tailor the course content with the topics teachers' acknowledge as relevant for their teaching. This premise is required for the success of the training course and to introduce new concepts into their

teaching, as has been shown by different studies (Form & Lewitter, 2011; Gallagher et al., 2011; Wood & Gebhardt, 2013).

A series of teacher training workshops have been promoted since 2010 by the European Learning Laboratory for the Life Sciences (ELLS) and European Bioinformatics Institute (EMBL-EBI). The main objectives of these training courses have been connecting bioinformatics with daily teaching practices, raising teachers' awareness about bioinformatics, and exploring new ways to visualize and explain biological concepts. This joint venture created a fruitful environment for the exchange of experiences between teachers from different countries (Wood & Gebhardt, 2013). As mentioned before, when designing a training course, it is extremely important to take into consideration teachers' needs and adequate proposals to the curricular contents, in order to smoothly integrate a new topic without disrupting the installed teaching routines by being an extra burden for teachers. The training interventions were targeted to teachers of secondary school level taking into consideration that it is at this stage that main genomic related topics are addressed, namely the DNA code, replication, transcription, and proteins' synthesis (Wood & Gebhardt, 2013). As in the TtGG training course, teachers were also actively involved in developing the activities which empower them with a sense of propriety.

The Netherlands Bioinformatics Centre (NBIC) together with the Freudenthal Institute (Netherlands Bioinformatics Centre, 2020, 2019a), are another example of research-driven institutions that support initiatives dedicated to teachers' training in bioinformatics. During the training, the teachers are challenged to use NAVIGENE (Netherlands Bioinformatics Centre, 2014) as a tool to provide a holistic perspective on genomics, genes and proteins through bioinformatics, and suitable for secondary education. These training programs are being developed as a part of a broader project to integrate bioinformatics in the classroom called "Bioinformatics@school" (Mil, 2007).

It is worth mentioning that a common feature of the mentioned teachers' training programs is to gather in-service teachers with distinct professional experiences or even from different countries. This networking is particularly important to strengthen teachers' confidence, to debate different perspectives, and to exchange successes and difficulties when exploring and implementing new bioinformatics-based approaches in their classes (LaRue et al., 2018). Furthermore, training courses fed teachers' sense of belonging to a community which is essential to generate *in situ* and post-course collaboration opportunities to mature solutions for pedagogical problems, or to re-designing their lessons or course materials making use of technology to improve learning (Palha, 2019).

Regardless the progress, it is interesting to notice that following the attendance of a training course, teachers are willing to deepen their knowledge in bioinformatics to make better use of the numerous resources available. In fact, even after several lessons in bioinformatics, online databases are still enigmatic to a majority of teachers. They are cautious to explore and prefer to know the direct path to an output rather than adopt an iterative or exploratory process of trial and error in database browsing (LaRue et al., 2018). This reality indicates that for teachers to feel comfortable to apply bioinformatics-based activities by themselves, they require a troubleshooter to hold on when needed, capable to further scaffold their learning (LaRue et al., 2018; Martins et al., 2020c; Wood & Gebhardt, 2013). In this regard, the TtGG program provides teachers with customized mobile laboratory kits, extensive curricula resources, and access to a 24/7 Genomics Help Desk (LaRue et al., 2018). A similar strategy is used by ELLS training course, in which an activity development session is promoted to provide an opportunity for participants to discuss how the proposed activities could be better adapted to their specific school reality and teaching environment (Wood & Gebhardt, 2013).

Prospectively, the dissemination of bioinformatics in schools cannot be limited to training in-service teachers. Careful attention should be given to the academic training of pre-service teachers through their master's programs in biology education (Porter & Smith, 2019). Furthermore, according to the main obstacles mentioned by teachers to not implement bioinformatics in their classrooms, namely the access to computers with a competent internet connection, and the time needed to adapt and translate the activities into a teaching language, cannot be neglected (Wood & Gebhardt, 2013). In fact, the most common external factors impeding technology integration in the classroom are the lack of access to computers and software, insufficient time to plan instruction, and inadequate technical and administrative support (Hew & Brush, 2007; Chen, 2008; Palha, 2019). It is, therefore, crucial to take into consideration schools' facilities and equipment to implement bioinformatics-based activities taking advantage of free-access tools.

Rationale and Objectives of this Study

The new paradigms of biological research demand the need to improve curricular and educational resources through initiatives validated by focused science education research. In this context, bioinformatics is at a privileged position to foster citizenship education. Taking this laboratory reality to the school has already proved to have a beneficial impact on students learning (Amenkhienan & Smith, 2006; Newman et al., 2016; Taylor et al., 2014). Moreover, activities that provide contact with resources used in research laboratories appear to be a stimulus to students' interest in scientific careers, allowing students to get acquainted with some daily routines of a professional researcher and get a glimpse of the training required to reach these expertise's (Kovarik et al., 2013). In this regard, it is urgent to engage educational stakeholders to foster bioinformatics education in schools and to support teachers by offering focused training courses that may allow teachers to confidently explore themes and tools with a considerable degree of demand. To meet this priority there are recent data showing that, after a first foray into bioinformatics activities, teachers tend to feel more motivated, confident and able to redesign the didactic-pedagogical goals that may embrace introductory bioinformatics (Attwood et al., 2017; Koch & Fuellen, 2008; Machluf, Gelbart, et al., 2017) (Figure I - 2).

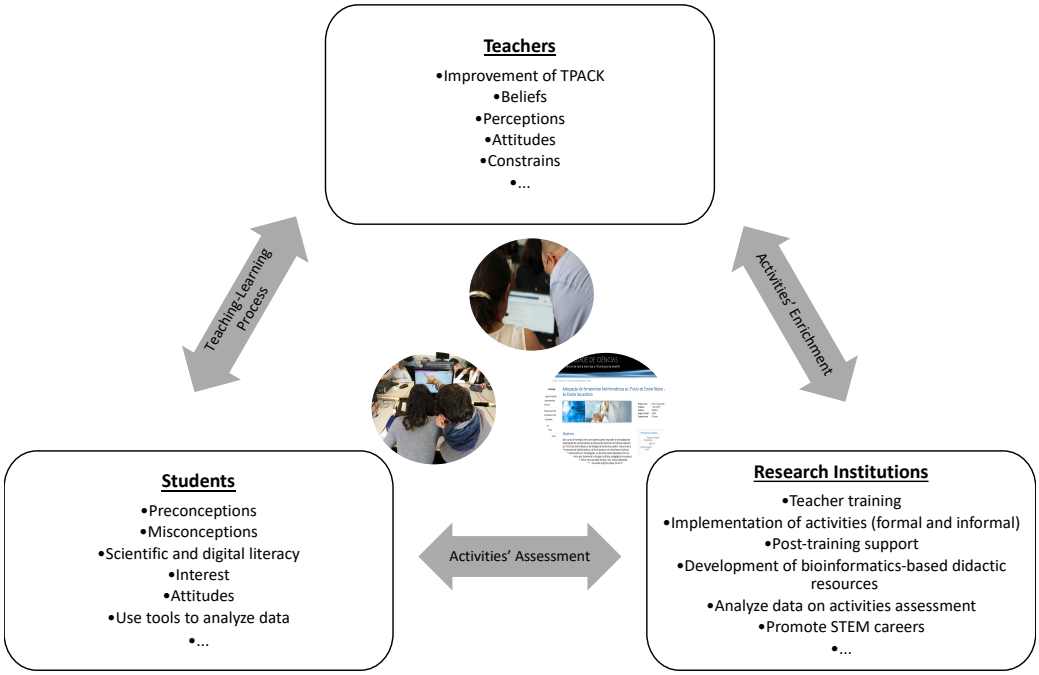


Figure I-2. Integration of bioinformatics as an educational resource entails a network of interactions between the stakeholders.

Despite the reported benefits about the integration of bioinformatics in elementary and secondary education, comprehensive assessments of the impact on students learning outcomes and teachers' practices were missing (Machluf, Gelbart, et al., 2017; Magana et al., 2014; Marques et al., 2014).

This doctoral project was aimed to select, adapt, implement and evaluate a set of bioinformatics resources framed within elementary and secondary curricular contexts in order to foster top-notch science education. The research approaches carried out are learner-centered and based on teaching-learning processes underlining the role of teachers as promoters of students' achievements.

Within the framework of this PhD project, the feature research questions were:

i) Are bioinformatics tools and resources decisive to foster student's scientific and digital literacy, interest and attitudes?

ii) Which are the main constraints that are preventing teachers to implement bioinformatics exercises in the classroom?

The project-specific aims were:

1. *The improvement, adjustment and implementation of bioinformatics-based activities as didactic resources.* To provide a portfolio of active learning-based bioinformatics activities curricular framed.
2. *The assessment of the impact on students' literacy, interest and attitudes towards bioinformatics in formal and non-formal educational contexts*. To evaluate the educational adequacy of different bioinformatics tools and resources, aiming to improve students' scientific and digital literacy, attitudes and interest, which are decisive to foster effective learning and accomplished citizenship education.
3. *The characterization of the perceptions and attitudes of teachers towards bioinformatics education.* This goal stems from the acknowledgment that although there are many suitable bioinformatics platforms to bring science topics to classrooms, teachers' limited knowledge in this field, and/or lack of time to select the fittest bioinformatics activities for a specific teaching objective, may prevent them from fully engaging in using bioinformatics resources for teaching.
4. *The assistance of teachers and the promotion of networking.* To conceive a dedicated webpage, to provide teachers with updated information and continuous support. Through this deliverable, the adapted bioinformatics tools and resources, namely guidelines, suggestions for curricular integration, and troubleshooting solutions, were made available. In addition, a communication

forum for teachers was offered, scaffolding collaborative work and networking. This was expected to support students learning, by providing learning resources and communication channels, so that the learning process could be continuously guided by teachers and institutions. Through diagnostic exercises, teachers may also monitor students' progress and identify difficulties, which is important to adjust the teaching practices.

The workflow of this study encompassed four main tasks:

Task 1 – Selection, Improvement, Adaptation and Implementation of Bioinformatics Activities

Within the framework of this doctoral project was required to select and adapt the most suitable bioinformatics tools to fit the Portuguese national curricula of Biology for elementary and secondary school levels. To boarder the implementation of the designed activities, Next Generation Science Standards (NGSS) (National Research Council, 2013) were also taken into consideration to contextualize the exercises proposed. In this context, a set of activities useful to teach specific themes and concepts were selected.

In order to improve and adjust these bioinformatics tools and resources, teachers' feedback was very important. Training courses for teachers and the design of a webpage website capable to provide continuous support, such as digital presentations and tutorials, and to gather updated information were the main deliverable of this task and considered essential to help teachers during the implementation. Teachers' collaboration was revealed to be critical to adjust the activities proposed, in order to meet both the instructors' and the students' needs and expectations and to fit the classes' schedule and the school computational resources.

Regarding the implementation of these activities, it was important to understand the benefits that they could bring to the students. Ultimately, this task provided learners with the opportunity to engage in authentic activities (i.e. like the ones carried out by scientists) and assisted teachers to address students' weaknesses and difficulties. The main endeavor of the task was to strengthen students' bioinformatics literacy, preparing them for future professional and academic challenges.

Due to its exhaustive nature of the curricula, and the time-constraints it imposes, this study focused on a restricted number of partner schools/classes.

The outputs and discussions around this task are detailed in **Chapter II** of this thesis.

Task 2 – Diagnose Students' Knowledge and Attitudes towards Bioinformatics

This task aimed to diagnose elementary (12-15 years old) and secondary school (15-18 years old) students' conceptualization of scientific topics such as genomics, genetics and evolution, and students' literacy, interest, and attitudes regarding bioinformatics and its integration in their learning practices.

The sampled universe comprised elementary and secondary school students, assessed in both formal and non-formal educational contexts. These instructional levels were purposely chosen to access students with different interests, scientific background and knowledge: while elementary school students are engaging in general training, which provides a good opportunity for them to get acquainted with bioinformatics tools and their potential, secondary school students are particularly interested in learning more about science, and expectably more motivated and enthusiastic to deal with innovative tools that can allow them to acquire the competencies pertained in their curriculum.

The educational adequacy of different bioinformatics activities was assessed by suitable instruments aiming to gather evidence about students' scientific and digital literacy, attitudes, and interest towards bioinformatics. A mixed-methods approach was used to characterize students' understanding and perceptions. Large-scale surveys were carried out to obtain statistically robust results.

The design, implementation and assessment of this task are detailed in **Chapter III** of this publication.

Task 3 – Perceptions and Attitudes of Teachers regarding Bioinformatics Education

The objective of this task was to depict the general landscape of teacher training, by characterizing teachers' academic background and identifying learning opportunities in the scope of bioinformatics. It is worth to recall that bioinformatics-related topics were not comprised in Pre-Bologna university curricula. With the transition to Post-Bologna degrees, bioinformatics was introduced in the curricular structure of some Masters' courses but was left out of pre-service teacher education. Therefore, biology teachers who are currently getting their academic degrees have poor training in bioinformatics, which may lead them to feel unconfident and misinformed to use bioinformatics tools in their classes.

To unveil teachers' perceptions about bioinformatics education and diagnose their acquired skills in basics bioinformatics, a specifically designed survey and non-formal interviews were performed, covering a range of topics, namely: academic background

and sense of proficiency; attitudes towards bioinformatics; and perceptions about the accessibility of bioinformatics resources. Teachers with different pedagogical experiences were sampled, ranging from those with an extensive academic experience to those who have just graduated or those who have upgraded their knowledge in bioinformatics.

The research developed under the goals of this task is detailed in **Chapter IV** of this publication.

Task 4 – Create a Webpage to Assist Teachers and Promote Networking

Framed within this task, a dedicated webpage was designed to provide teachers with updated information and support.

Besides displaying the adapted bioinformatics tools and resources, this webpage includes guidelines for teachers, suggestions for curricular integration, and troubleshooting solutions, to empower the teachers to implement the chosen activities. Furthermore, this webpage includes a forum to strengthen teachers' networking, in which they are encouraged to share ideas for new activities, remodeling of others, or add relevant information (e.g. students' difficulties, institutional constraints, curricular framing, or pedagogic counselling).

Considerations on the webpage design, sharing and use by teachers are described in **Chapter V**.

Thesis Publications

The major part of the results gathered during this doctoral project were published in international peer-reviewed journals and conference proceedings of international conferences.

The publications have been edited for this thesis to ensure a coherent format, although the contents are exactly the same as those published.

Publications in international peer-reviewed journals:

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher, 80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619 **(Chapter II)**

Martins, A., Fonseca, M. J., Lemos, M., Lencastre, L., & Tavares, F. (2020). Bioinformatics-Based Activities in High School: Fostering Students' Literacy, Interest and Attitudes on Gene Regulation, Genomics and Evolution. *Frontiers in Microbiology, 11,* 578099. https://doi.org/10.3389/fmicb.2020.578099 **(Chapter III)**

Publications in international conference proceedings:

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network. ISBN: 978-84-8158-779-1 **(Chapter II)**

Martins, A., Lencastre, L., & Tavares, F. (2018). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education*. (pp. 309–317). Hands-on Science Network. ISBN: 978-84-8158-779-1 **(Chapter III)**

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of Bioinformatics Tools to Elementary and Secondary School Curricula: a Training Course for Teachers. In *Congresso Internacional - O Tempo dos Professores.* (pp. 515-522). FPCEUP. ISBN: 978-989-8471-26-0 **(Chapter IV)**

Martins, A., Lencastre, L., & Tavares, F. (2018). Integrating Bioinformatics in Elementary and Secondary Education: Teacher's Perceptions. In *3rd International*

Conference on Teacher Education (INCTE). (pp. 203-2014). Instituto Politécnico de Bragança. ISBN: 978-972-745-241-5 **(Chapter IV)**

Martins, A., Lencastre, L., & Tavares, F. (2020). Bioinformatics, a Befitting Tool for e-Learning: Potential and Constraints according Teachers' Perceptions. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.). *Hands-on Science. Science Education. Discovering and understanding the wonders of Nature*. (pp. 97-105). Hands-on Science Network. ISBN: 978-84-8158-841-5 **(Chapter IV)**

Martins, A., Lencastre, L., & Tavares, F. (2020). Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers. *The Beauty and Pleasure of Understanding: Engaging with Contemporary Challenges Through Science Education*. (pp. 1712-1721). European Science Education Research Association. ISBN: 978-88-945874-0-1 **(Chapter IV)**

Martins, A., Lencastre, L., & Tavares, F. (2020). "Bioinformática na Sala de Aula": Webpage to Boost Bioinformatics in the Classroom. *Simpósio Internacional de Psicologia Da Educação: Passado, Presente e Futuro* (SInPE20). **(Chapter V)**

References

23andMe. (2020). *DNA Genetic Testing & Analysis - 23andMe*. Retrieved January 16, 2020, from https://www.23andme.com/

Adam, T. (2020). Personalized and Precision Medicine Informatics Education. *Personalized and Precision Medicine Informatics* (319-330). Springer, Cham. https://doi.org/10.1007/978-3-030-18626-5_20

Alaie, A., Teller, V., & Qiu, W. (2012). A Bioinformatics Module for Use in an Introductory Biology Laboratory. *The American Biology Teacher*, *74*(5), 318–322. https://doi.org/10.1525/abt.2012.74.5.6

Altman, R. (2009). *Bioinformatics & Computational Biology = same? No.* Building Confidence. https://rbaltman.wordpress.com/2009/02/18/bioinformatics-computational-biology-same-no/

Altschul, S., Gish, W., Miller, W., Myers, E., & Lipman, D. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amenkhienan, E., & Smith, E. (2006). A web-based genetic polymorphism learning approach for high school students and science teachers. *Biochemistry and Molecular Biology Education*, *34*(1), 30–33. https://doi.org/10.1002/bmb.2006.49403401030

Anderson, N., Ash, J., & Tarczy-Hornoch, P. (2007). A qualitative study of the implementation of a bioinformatics tool in a biological research laboratory. *International Journal of Medical Informatics*, *76*(11–12), 821–828. https://doi.org/10.1016/j.ijmedinf.2006.09.022

Athanasiadis, G., Jørgensen, F., Cheng, J., Kjærgaard, P., Schierup, M., & Mailund, T. (2016). Spitting for Science: Danish High School Students Commit to a Large-Scale Self-Reported Genetic Study. *PLoS ONE*, *11*(8), e0161822. https://doi.org/10.1371/journal.pone.0161822

Attwood, T., Blackford, S., Brazas, M., Davies, A., & Schneider, M. (2017). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbx100

Bacusmo, J., Bokor, J., Savage, K., & de Crécy-Lagard, V. (2019). Identifying Pathogenic Islands through Genome Comparison. *The American Biology Teacher*, *81*(8), 577–581. https://doi.org/10.1525/abt.2019.81.8.577

Bahar, M., Johnstone, A., & Hansell, M. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, *33*(2), 84–86. https://doi.org/10.1080/00219266.1999.9655648

Barcelona Supercomputing Center. (2019). *Bioinfo 4 women – Outstanding Young Female Bioinformaticians*. http://bioinfo4women.bsc.es/

Belmont, J., & Shaw, C. (2016). Clinical bioinformatics: emergence of a new laboratory discipline. *Expert Review of Molecular Diagnostics*, *16*(11), 1139–1141. https://doi.org/10.1080/14737159.2016.1246184

Boaler, J., & Levitt, S. (2019). *Opinion: High school math should be about data science, not Algebra 2 - Los Angeles Times*. https://www.latimes.com/opinion/story/2019-10-23/math-high-school-algebra-data-statistics

Bokor, J., Landis, J., & Crippen, K. (2014). High school students' learning and perceptions of phylogenetics of flowering plants. *CBE Life Sciences Education*. https://doi.org/10.1187/cbe.14-04-0074

Børne - og undervisningsministeriet. (2017a). *Biologi A – stx, august 2017*. Stx - Læreplaner 2017. https://www.uvm.dk/gymnasiale-uddannelser/fag-og-laereplaner/laereplaner-2017/stx-laereplaner-2017

Børne - og undervisningsministeriet. (2017b). Bioteknologi A – stx, august 2017. *Stx - Læreplaner 2017*.

Boyd, C. (2020). *23andMe inks deal with drug firm to develop antibody for skin conditions | Daily Mail Online*. https://www.dailymail.co.uk/health/article-7872491/23andMe-inks-deal-drug-firm-develop-antibody-skin-conditions.html

Brazas, M., Lewitter, F., Schneider, M., van Gelder, C., & Palagi, P. (2014). A Quick Guide to Genomics and Bioinformatics Training for Clinical and Public Audiences. *PLoS Computational Biology*, *10*(4). https://doi.org/10.1371/journal.pcbi.1003510

Bundesministerium für Bildung Wissenschaft und Forschung. (2019). *RIS - BGBLA_2019_II_107 - Bundesgesetzblatt authentisch ab 2004*. https://www.ris.bka.gv.at/eli/bgbl/II/2019/107

Campbell, C., & Nehm, R. (2013). A critical analysis of assessment quality in genomics and bioinformatics education research. *CBE Life Sciences Education*, *12*(3), 530–541. https://doi.org/10.1187/cbe.12-06-0073

Cham, J. (2013). *ISMB/ECCB Conferences.*

https://www.flickr.com/photos/97823772@N02/sets/72157634564701268/with/9367535935/

Chen, C. (2008). Why do teachers not practice what they believe regarding technology integration? *Journal of Educational Research*, *102*(1), 65–75. https://doi.org/10.3200/JOER.102.1.65-75

Chiovitti, A., Thorpe, F., Gorman, C., Cuxson, J., Robevska, G., Szwed, C., Duncan, J., Vanyai, H., Cross, J., Siemering, K., & Sumner, J. (2019). A citizen science model for implementing statewide educational DNA barcoding. *PLoS ONE, 14*(1), e0208604. https://doi.org/10.1371/journal.pone.0208604

Coccia, M. (2020). Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. *Technology in Society*, *60*. https://doi.org/10.1016/j.techsoc.2019.101198

Collins, F., Green, E., Guttmacher, A., & Guyer, M. (2003). A vision for the future of genomics research. *Nature*, *422*(6934), 835–847. https://doi.org/10.1038/nature01626

Cummings, M., & Temple, G. (2010). Broader incorporation of bioinformatics in education: Opportunities and challenges. *Briefings in Bioinformatics*, *11*(6), 537–543. https://doi.org/10.1093/bib/bbq058

David, A. (2018). Using Project-Based Learning to Teach Phylogenetic Reconstruction for Advanced Undergraduate Biology Students: Molluscan Evolution as a Case Study. *The American Biology Teacher*, *80*(4), 278–284. https://doi.org/10.1525/abt.2018.80.4.278

Department for Education (UK Government). (2014). *Science programmes of study: key stage 4 National curriculum in England*.

Dressler, L., Jones, S., Markey, J., Byerly, K., & Roberts, M. (2014). Genomics education for the public: Perspectives of genomic researchers and ELSI advisors. *Genetic Testing and Molecular Biomarkers*, *18*(3), 131–140. https://doi.org/10.1089/gtmb.2013.0366

Dubey, R., Tripathi, V., Prabha, R., Chaurasia, R., Singh, D., Rao, C., El-Keblawy, A., & Abhilash, P. (2020). *Bioinformatics Tools for Soil Microbiome Analysis* (61–70). https://doi.org/10.1007/978-3-030-15516-2_6

Duffus, A. (2019). An Emerging Amphibian Infection as a Model for Teaching Phylogenetic Reconstruction. *The American Biology Teacher*, *81*(1), 32–39. https://doi.org/10.1525/abt.2019.81.1.32

Duncan, R., Rogat, A., & Yarden, A. (2009). A learning progression for deepening students' understandings of modern genetics across the 5th-10th grades. *Journal of Research in Science Teaching*, *46*(6), 655–674. https://doi.org/10.1002/tea.20312

Eble, J., & Pecore, J. (2019). "Invasive Aliens": A Student Citizen-Science Activity Using DNA Barcoding to Investigate Concepts in Ecology & Molecular Biology. *The American Biology Teacher*, *81*(3), 169–174. https://doi.org/10.1525/abt.2019.81.3.169

Einheitliche Prüfungsanforderungen in der Abiturprüfung Biologie. (2004).

Eurich, C., Fields, P., & Rice, E. (2012). Proteomics: Protein Identification Using Online Databases. *The American Biology Teacher*, *74*(4), 250–255. https://doi.org/10.1525/abt.2012.74.4.8

European Bioinformatics Institute. (2019). *Training | European Bioinformatics Institute*. https://www.ebi.ac.uk/training

Fernandes, C. (2019). *A metadata analysis of walnut associated Xanthomonas spp.: from population diversity to comparative genomics* [University of Porto]. https://hdl.handle.net/10216/122012

Fernandes, E., Dias, C., Fonseca, M. J., & Tavares, F. (2014). Understanding Growth and Thermal Inactivation of Foodborne Bacteria Using the Pathogen Modelling Program (PMP). In Manuel Costa, P. Pombo, & J. Dorrio (Eds.), *Hands-on Science: Science Education with and for Society* (pp. 207–210). Hands-on Science Network.

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. *PLoS Computational Biology* *7*(10), e1002243. https://doi.org/10.1371/journal.pcbi.1002243

Fundação Calouste Gulbenkian - Instituto Gulbenkian de Ciência. (2012). *Bioinformática na Escola*. http://en.bioinformatica-na-escola.org/

Gallagher, S., Coon, W., Donley, K., Scott, A., & Goldberg, D. (2011). A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Computational Biology*, *7*(10). https://doi.org/10.1371/journal.pcbi.1002244

Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research simulation in bioinformatics. *Journal of Biological Education*, *40*(3), 107–112. https://doi.org/10.1080/00219266.2006.9656026

Gurwitz, D., Weizman, A., & Rehavi, M. (2003). Education: Teaching pharmacogenomics to prepare future physicians and researchers for personalized medicine. *Trends in Pharmacological Sciences,* *24*(3), 122–125).

https://doi.org/10.1016/S0165-6147(03)00024-5

Hamilton, A. (2008). *1. The Retail DNA Test - Best Inventions of 2008 - TIME*. http://content.time.com/time/specials/packages/article/0,28804,1852747_1854493_1854113,00.html

Hamzelou, J. (2020). *23andMe has sold the rights to develop a drug based on its users' DNA | New Scientist*. https://www.newscientist.com/article/2229828-23andme-has-sold-the-rights-to-develop-a-drug-based-on-its-users-dna/

Haraldsrud, A. (2017). *Programmering og modellering - Valler videregående skole*. https://www.valler.vgs.no/aktuelt/nyhetsarkiv/?article_id=95139

Hesper, B., & Hogeweg, P. (1970). Bioinformatica: een werkconcept. *Kameleon*, *1*(6), 28–29.

Hew, K., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, *55*(3), 223–252. https://doi.org/10.1007/s11423-006-9022-5

Hogeweg, P. (2011). The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*, *7*(3), e1002021. https://doi.org/10.1371/journal.pcbi.1002021

Huerta, M., Downing, G., Haseltine, F., Seto, B., & Liu, Y. (2000). *NIH Working Definition of Bioinformatics and Computational Biology*. https://2digitstechcom.ipage.com/uploads/2/9/0/1/2901227/compubiodef.pdf

Husum, P., & Haraldsrud, A. (2017). *Forsøkslæreplan i programmering og modellering X - programfag i utdanningsprogram for studiespesialisering (PRM1-01)*. https://www.udir.no/kl06/PRM1-01

Iannetti, L., Salini, R., Sperandii, A., Santarelli, G., Neri, D., Di Marzio, V., Romantini, R., Migliorati, G., & Baranyi, J. (2017). Predicting the kinetics of Listeria monocytogenes and Yersinia enterocolitica under dynamic growth/death-inducing conditions, in Italian style fresh sausage. *International Journal of Food Microbiology*, *240*, 108–114. https://doi.org/10.1016/j.ijfoodmicro.2016.04.026

International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, *409*(6822), 860–921. https://doi.org/10.1038/35057062

Juran, B., & Server, A. (2001). *Bioinformatics - New Legal Issues Raised in the Intersection of Computer Science, Biology and the Internet | WilmerHale*.

https://www.wilmerhale.com/en/insights/publications/bioinformatics-new-legal-issues-raised-in-the-intersection-of-computer-science-biology-and-the-internet-november-5-2001

Kawrykow, A., Roumanis, G., Kam, A., Kwak, D., Leung, C., Wu, C., Zarour, E., Sarmenta, L., Blanchette, M., & Waldispühl, J. (2012). Phylo: A Citizen Science Approach for Improving Multiple Sequence Alignment. *PLoS ONE, 7*(3), e31362. https://doi.org/10.1371/journal.pone.0031362

Kelling, S. (2012). Using Bioinformatics in Citizen Science. In J. Dickinson & R. Bonney (Eds.), *Citizen Science: Public Participation in Environmental Research* (1st ed.*)*. Cornell University Press.

Kesh, S., & Raghupathi, W. (2004). Critical issues in bioinformatics and computing. *Perspectives in Health Information Management*, *1*(9). http://www.ncbi.nlm.nih.gov/pubmed/18066389

Koch, I., & Fuellen, G. (2008). A review of bioinformatics education in Germany. *Briefings in Bioinformatics*, *9*(3), 232–242. https://doi.org/10.1093/bib/bbn006

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Kovarik, D., Patterson, D., Cohen, C., Sanders, E., Peterson, K., Porter, S., & Chowning, J. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Lafta, I. (2020). Bioinformatics and molecular analysis of the breast cancer susceptibility gene brca1 in breast cancer. *Middle East Journal of Cancer*, *11*(1), 59–71. https://doi.org/10.30476/mejc.2019.83434.1165

LaRue, K., McKernan, Mm., Bass, K., & Wray, C. (2018). Teaching the Genome Generation: Bringing Modern Human Genetics into the Classroom Through Teacher Professional Development. *The Journal of STEM Outreach*, *1*(1). https://doi.org/10.15695/jstem/v1i1.12

Lesk, A. (2019). *Introduction to bioinformatics*. Oxford University Press.

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: re-visiting students' understandings of genetics. *International Journal of Science Education*, *26*(2), 195–206. https://doi.org/10.1080/0950069032000072782

Lewitter, F., & Bourne, P. (2011). Teaching Bioinformatics at the Secondary School

Level. *PLoS Computational Biology*, *7*(10), e1002242. https://doi.org/10.1371/journal.pcbi.1002242

Luscombe, N., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*. https://doi.org/10.1053/j.ro.2009.03.010

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Machluf, Y., Tal, O., Navon, A., & Chaiter, Y. (2017). From Population Databases to Research and Informed Health Decisions and Policy. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00230

Magana, A., Taleyarkhan, M., Alvarado, D., Kane, M., Springer, J., & Clase, K. (2014). A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE—Life Sciences Education*, *13*(4), 607–623. https://doi.org/10.1187/cbe.13-10-0193

Marques, I., Almeida, P., Alves, R., Dias, M., Godinho, A., & Pereira-Leal, J. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r1_id=

Martins, A., Lencastre, L., & Tavares, F. (2018a). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. In *3rd International Conference on Teacher Education (INCTE)*. Instituto Politécnico de Bragança. http://hdl.handle.net/10198/17381

Martins, A., Lencastre, L., & Tavares, F. (2018b). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 309–317). Hands-on Science Network.

Martins, A, Fonseca, M., Lemos, M., Lencastre, L., & Tavares, F. (2020a). Bioinformatics-based activities in high school: Fostering students' literacy, interest and attitudes on gene regulation, genomics and evolution. *Frontiers in Microbiology, 11,* 578099. https://doi.org/10.3389/fmicb.2020.578099

Martins, A., Lencastre, L., & Tavares, F. (2020b). Bioinformatics, a Befitting Tool for e-learning: Potential and Constraints according Teachers' Perceptions. In M. F. Costa & J. B. Dorrío (Eds.), *Hands-on Science. Science Education. Discovering and understanding the wonders of Nature* (pp. 97–105). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2020c). Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers. *The Beauty and Pleasure of Understanding: Enganging with Contemporary Challenges Through Science Education*. European Science Education Research Association. *In Press.*

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In Manuel Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

Mathur, M. (2018). Bioinformatics challenges: A review. *International Journal of Advanced Scientific Research*, *3*(6), 29–33. https://www.researchgate.net/publication/329529285

Mendes, A., Rebelo, D., & Pinheiro, E. (2003). *Programa de Biologia e Geologia 11º ou 12ºano(s)*. Ministério da Educação: Departamento do Ensino Secundário.

Mendes, A., Rebelo, D., & Pinheiro, E. (2004). *Programa Curricular Biologia 12ºano*.

Merelli, I., Pérez-Sánchez, H., Gesing, S., & D'Agostino, D. (2014). Managing, Analysing, and Integrating Big Data in Medical Bioinformatics: Open Problems and Future Perspectives. *BioMed Research International*, *2014*. https://doi.org/10.1155/2014/134023

Mil, M. (2007). DNA labs on the road. *Science in School*. https://www.scienceinschool.org/2007/issue6/dnalabs

Ministério da Educação. (2018a). *Aprendizagens Esseciais - Biologia 12ºano*.

Ministério da Educação. (2018b). *Aprendizagens Essenciais - Biologia e Geologia 11º ano*.

Moore, A. (2008). Science teaching must evolve. *Nature*, *453*(7191), 31–32. https://doi.org/10.1038/453031a

Mumtaz, S. (2000). Factors affecting teachers' use of information and communications technology: a review of the literature. *Journal of Information Technology for Teacher Education*, *9*(3), 319–342. https://doi.org/10.1080/14759390000200096

Murphy, R. (n.d.). *What is Computational Biology? | Computational Biology Department*. Retrieved October 16, 2019, from http://www.cbd.cmu.edu/about-us/what-is-computational-biology/

National Aeronautics and Space Administration. (2019). *GeneLab for High Schools: Growing the Next Generation of Scientists | NASA*. https://www.nasa.gov/ames/genelab-for-high-schools

National Geographic. (2016). *The Genographic Project®* . Geno 2.0 Next Generation Helix Product Privacy Policy. https://www.nationalgeographic.com/legal/privacy/genographic/

National Human Genome Research Institute. (2019). *The Human Genome Project*. https://www.genome.gov/human-genome-project

National Research Council. (2012). Bioinformatics: Interpreting the Human Genome. In *Fueling Innovation and Discovery: The Mathematical Sciences in the 21st Century* (pp. 44–48). National Academies Press. https://doi.org/10.17226/13373

National Research Council. (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

nature.com. (2020). *Open reading frames - Latest research and news | Nature*. https://www.nature.com/subjects/open-reading-frames

NCBI. (2020a). *BLAST: Basic Local Alignment Search Tool*. https://blast.ncbi.nlm.nih.gov/Blast.cgi

NCBI. (2020b). *Home - ORFfinder - NCBI*. https://www.ncbi.nlm.nih.gov/orffinder/

Netherlands Bioinformatics Centre. (2020). *NBIC: Teacher training*. Retrieved February 13, 2020, from https://www.nbic.nl/education/high-school-programmes/bioinformaticsschool/teacher-training/index.html

Netherlands Bioinformatics Centre. (2014). *NBIC: NAVIGENE*. https://www.nbic.nl/education/high-school-programmes/bioinformaticsschool/teacher-

training/navigene/

Netherlands Bioinformatics Centre. (2019a). *NBIC: Bioinformatics@school*. https://www.nbic.nl/education/high-school-programmes/bioinformaticsschool/index.html

Netherlands Bioinformatics Centre. (2019b). *NBIC: BioWise*. https://www.nbic.nl/education/biowise/index.html

Newman, L., Duffus, A., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

NGSS. (2013). *HS-LS3-3 Heredity: Inheritance and Variation of Traits | Next Generation Science Standards*. https://www.nextgenscience.org/pe/hs-ls3-3-heredity-inheritance-and-variation-traits

Ouzounis, C. (2012). Rise and demise of bioinformatics? promise and progress. *PLoS Computational Biology*, *8*(4). https://doi.org/10.1371/journal.pcbi.1002487

Palha, S. (2019). Student teachers´ needs and preferences for a technology preparation course. In L. Leite, E. Oldham, L. Caravalho, A. Afonso, F. Viseu, L. Dourado, & M. Martinho (Eds.), *ATEE Winter Conference 2019 – Science and mathematics education in the 21st century* (pp. 72–81). ATEE and CIEd.

Parsons, J. (2002). Editorial: Bioinformatics and economics. *Briefings in Bioinformatics*, *3*(4), 328–330. https://doi.org/10.1093/bib/3.4.328

Patrinos, G. (2020). Applied Genomics and Public Health. In *Applied Genomics and Public Health* (pp. 1–7). Elsevier. https://doi.org/10.1016/b978-0-12-813695-9.00001-7

Pirinen, I. (2019). *Bioinformatiikan hyödyntäminen lukion biologian opetuksessa* [Itä-Suomen yliopisto]. http://urn.fi/urn:nbn:fi:uef-20191015

Porter, S., & Smith, T. (2019). Bioinformatics for the Masses: The Need for Practical Data Science in Undergraduate Biology. *Omics : A Journal of Integrative Biology*, *23*(6), 297–299. https://doi.org/10.1089/omi.2019.0080

Rahn, J., Willner, D., Deverick, J., Kemper, P., & Saha, M. (2019). Incorporating Computer Programming & Data Science into a Guided Inquiry-Based Undergraduate Ecology Lab. *The American Biology Teacher*, *81*(9), 649–657. https://doi.org/10.1525/abt.2019.81.9.649

Sadek, H. (2004). *Bioinformatics : principles, basic internet applications*. Trafford Publishing.

Sayres, W., Hauser, C., Sierk, M., Robic, S., Rosenwald, A., et al.

(2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS ONE, 13*(6): e0196878. https://doi.org/10.1371/journal.pone.0196878

Shachak, A., & Fine, S. (2008). The effect of training on biologists acceptance of bioinformatics tools: A field experiment. *Journal of the American Society for Information Science and Technology*, *59*(5), 719–730. https://doi.org/10.1002/asi.20772

Shachak, A., Shuval, K., & Fine, S. (2007). Barriers and enablers to the acceptance of bioinformatics tools: A qualitative study. *Journal of the Medical Library Association*, *95*(4), 454–458. https://doi.org/10.3163/1536-5050.95.4.454

Stern, F., & Kampourakis, K. (2017). Teaching for genetics literacy in the post-genomic era. *Studies in Science Education*, *53*(2), 193–225. Routledge. https://doi.org/10.1080/03057267.2017.1392731

Swiss Institute of Bioinformatics. (2019). *SIB Training - Who can benefit*. https://www.sib.swiss/training/who-can-benefit

Szarecka, A., & Dobson, C. (2019). Protein Structure Analysis: Introducing Students to Rational Drug Design. *The American Biology Teacher*, *81*(6), 423–429. https://doi.org/10.1525/abt.2019.81.6.423

Taylor, D., Campbell, A., & Heyer, L. (2013). Illuminating the Black Box of Genome Sequence Assembly. *The American Biology Teacher*, *75*(8), 572–577. https://doi.org/10.1525/abt.2013.75.8.9

Taylor, J., Davidson, R., & Strong, M. (2014). Drug-resistant tuberculosis: A genetic analysis using online bioinformatics tools. *American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Thorn, C., Klein, T., & Altman, R. (2010). Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics*, *11*(4), 501–505. https://doi.org/10.2217/pgs.10.15

Undervisningsministeriet, & Styrelsen for Undervisning og Kvalitet. (2018a). *Biologi A/B/C, stx Vejledning*.

Undervisningsministeriet, & Styrelsen for Undervisning og Kvalitet. (2018b). *Bioteknologi A, stx Vejledning*.

University of Cambridge. (2019). *Bioinformatics Training*. https://bioinfotraining.bio.cam.ac.uk/

Valenzuela, A., Mullins, E., & Lauer, A. (2019). Detecting a Fungal Pathogen in Its Natural Habitat: The Case of Valley Fever. *The American Biology Teacher*, *81*(7), 492–501. https://doi.org/10.1525/abt.2019.81.7.492

Van Gelder, C., Hooft, R.., Van Rijswijk, M., Van Den Berg, L., Kok, R., Reinders, M., Mons, B., & Heringa, J. (2019). Bioinformatics in the Netherlands: The value of a nationwide community. *Briefings in Bioinformatics*, *20*(2), 375–383. https://doi.org/10.1093/bib/bbx087

Vereniging NLT. (2014). *Vereniging NLT - natuur leven technologie*. https://betavak-nlt.nl/nl/p/english/

Voorhuijzen-Harink, M., Hagelaar, R., van Dijk, J., Prins, T., Kok, E., & Staats, M. (2019). Toward on-site food authentication using nanopore sequencing. *Food Chemistry: X*, *2*, 100035. https://doi.org/10.1016/j.fochx.2019.100035

Wageningen University. (2014). *Bioinformatica: moleculaire biologie achter de computer*. http://vaklokaal-nlt.nl/?p=55

Wefer, S. (2003). Name That Gene: An Authentic Classroom Activity Incorporating Bioinformatics. *The American Biology Teacher*, *65*(8), 610–613. https://doi.org/10.2307/4451571

Wefer, S., & Anderson, O. (2008). Identification of Students' Content Mastery and Cognitive and Affective Percepts of a Bioinformatics Miniunit: A Case Study With Recommendations for Teacher Education. *Journal of Science Teacher Education*, *19*(4), 355–373. https://doi.org/10.1007/s10972-008-9099-2

Wefer, S., & Sheppard, K. (2008). Bioinformatics in High School Biology Curricula: A Study of State Science Standards. *CBE—Life Sciences Education*, *7*(1), 155–162. https://doi.org/10.1187/cbe.07-05-0026

Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., & Schneider, M. (2014). Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. *PLoS Computational Biology*, *10*(3), e1003496. https://doi.org/10.1371/journal.pcbi.1003496

Whitley, K., Tueller, J., & Weber, K. (2020). Genomics Education in the Era of Personal Genomics: Academic, Professional, and Public Considerations. *International Journal of Molecular Sciences*, *21*(3), 768. https://doi.org/10.3390/ijms21030768

Wightman, B., & Hark, A. (2012). Integration of bioinformatics into an undergraduate biology curriculum and the impact on development of mathematical skills. *Biochemistry and Molecular Biology Education*, *40*(5), 310–319. https://doi.org/10.1002/bmb.20637

Wood-Robinson, C., Lewis, J., & Leach, J. (2000). Young people's understanding of the nature of genetic information in the cells of an organism. *Journal of Biological Education*, *35*(1), 29–36. https://doi.org/10.1080/00219266.2000.9655732

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School-New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6). https://doi.org/10.1371/journal.pcbi.1003089

Wray, C. (2017). Introducing Students to the Genome: Brave New World or the Red Queen's Wonderland? *The American Biology Teacher*, *79*(4), 253–253. https://doi.org/10.1525/abt.2017.79.4.253

Yang, X., Hartman, M. R., Harrington, K. T., Etson, C. M., Fierman, M. B., Slonim, D. K., & Walt, D. R. (2017). Using Next-Generation Sequencing to Explore Genetics and Race in the High School Classroom. *CBE—Life Sciences Education*, *16*(2), ar22. https://doi.org/10.1187/cbe.16-09-0281

Chapter II

Bioinformatics-based Activities: Improvement, Adjustment and Implementation

Chapter II includes the following publications:

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network. ISBN: 978-84-8158-779-1

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher, 80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Genomics Education: Update Core Concepts in High School

Abstract. *Genomics is benefiting from exciting technological advances, joining next generation sequencing platforms with bioinformatics solutions for data analysis. This paradigm cannot be disregarded when promoting genomic literacy among high school students. In this work we explored a bioinformatics-based approach designed to have a positive impact on students' knowledge about genomics' concepts and methods. Based on it, this study highlights the benefits and the adequacy of dedicated bioinformatics activities to foster students learning of core concepts in genomics, while contributing to enhance students' motivation, interest and scientific reasoning.*

Keywords. Bioinformatics, Conceptions, Genomics, High School.

1. Introduction

Genomics can be briefly defined as a scientific field dedicated to the study of genomes (EMBL-EBI, 2018). In the last decade, it has had major developments, mainly due to technological advances of next generation sequencing platforms coupled with ingenious bioinformatics solutions for data analysis (Koboldt et al., 2013; Mardis, 2008; Schuster, 2008).

In this regard, it is widely acknowledged by educational stakeholders that genomics literacy, briefly defined as the knowledge of basic genetics and genomic concepts and processes, should be promoted in high school (Eijck, 2010; Hurle et al., 2013).

The integration of basic genomics concepts in the classroom should incorporate the use of bioinformatics and computational biology tools having in mind the importance of these resources to comprehensively address genomics studies (Ditty et al., 2010; Lewitter & Bourne, 2011; McQueen et al., 2012; Wefer & Sheppard, 2008).

The interdisciplinary character of bioinformatics is at a privileged position to foster citizenship education (Marques et al., 2014). A bioinformatics-based approach increases students' awareness of the decisive role of other disciplines, such as mathematics - through algorithms for data analysis - and computer science – capable to integrate large datasets - in genomics research (Table II - 1).

Table II-1. Potential of bioinformatics tools as a promoter of interdisciplinarity: in Information Technologies and Communication these tools can be used to understand the role of information technology in science development; in Chemistry, teachers can use these resources to manipulate the factors that influence chemical reactions; or in Mathematics as a source of graphs for students to interpret.

	Curricular framework	Learning Goals
Information and Communication Technology	• Information, knowledge and the world of technology: the evolution of information and communication technologies (ICT) and its role in the contemporary world. • Exploration of computing environments: creation of products, using tools and computing environments installed locally or available on the Internet, appropriate to the cognitive development of students.	• To develop knowledge and skills in the use of information and communication technologies that allow widespread digital literacy. • To foster the critical analysis of the role and power of information and communication technologies. • To develop a method of computational thinking, centered in the description and problem solving and in the logical organization of ideas. • To stimulate students as active users of computers, networks and Internet.
Chemistry	• Physical and chemical properties of materials: physical and chemical properties of substances. Justify from selected information, the importance of analytical chemistry in areas related to our quality of life, such as food security, environmental quality and disease diagnosis. • Chemistry and Industry: Control of industrial production; Effects of temperature and concentration in the equilibrium of a system.	• To understand that the manipulation of chemical reactions can promote a desired outcome. • To disclose the influence of factors that generally affect the equilibrium in a system, and their relevance in everyday situations, such as in the food industry.
Mathematics	• Measures of location: Represent, process and analyze data sets. • Solving problems involving functions and manipulating variables.	• To organize, analyze and design solutions for a problem by interpreting data displayed graphically.

Genomics education is fundamental to build up informed students capable to engage judiciously into discussions about genomics solutions for a panoply of real-world problems (e.g. gene therapies, genetically modified organisms – GMOs -, cloning, genetic testing or genetically engineered vaccines) (Črne-Hladnik et al., 2009; Dawson, 2007; Massarani & De Castro Moreira, 2005; Munn et al., 1999; Sadler & Zeidler, 2004). According to Kovarik *et al.* (Kovarik et al., 2013), discussing real-world problems with students and introducing them to the exploration of authentic science tools, contributes to increase their interest in Science, Technology, Engineering and Mathematics (STEM) contents, while promoting critical thinking. Moreover, these approaches involve ethical theory which help students to understand the relevance of the science, leading them to assume a position regarding society issues with

impact in our daily lives. In fact, in the last decade we witnessed a burst of exciting findings and societal discussions related, for instance, with personalized medicine based on individual genomic information, preservation of biodiversity, and the promise of new molecules from comprehensive metagenomics studies. These scientific progresses are raising sensible questions regarding bioethics and political options, that will require educated citizens capable to take scientifically informed decisions.

By getting acquainted with bioinformatics, students will realize their applicability and foster their interest, which may ultimately contribute to pursue careers in STEM fields. For students who do not pursue careers in STEM, understanding the applications and limitations of bioinformatics tools will scaffold them in taking informed decisions (Kovarik et al., 2013).

Regardless the importance of this subject, schoolteachers generally feel uneasy to approach this issue and tend to centre their teaching practice in expository methods of manual contents (Martins et al., 2017; Martins, Lencastre, et al., 2018). In this regard, it is urgently needed to propose hands-on bioinformatics-based activities aiming to introduce a practical component to boost the learning outcomes (Abrahams & Millar, 2008; Kibirige et al., 2014). To fully acknowledge the importance of the currently curricular required notions, an update of the science standards of high school education is required in order to include new core concepts to face the challenges of an era characterized by daily advances in genomics and metagenomics.

2. Genomics in High School Curriculum

Nowadays, basic notions of genomics already integrate the high school science curriculum to address diverse issues, particularly those related with heredity, biological evolution, gene regulation and protein synthesis (National Research Council, 2013; Mendes et al., 2003). Some examples of the required notions high school students have to acquire are listed in Table II - 2.

Table II-2. Description of required notions which currently integrate the science standards for high school; and of the core concepts which would be important to add in the curriculum to improve scientific literacy in genomics.

Improving Genomics Education *Genomics Concepts*	
Current Required Notions	**Introducing New Core Concepts**
Genome Chromosomes Genes (structural, operator, repressor, regulator, promotor) Start and stop codons Operons Genetic code Taxonomic groups Evolutionary relations	Open reading frames (ORFs) Basic Local Alignment Tool (BLAST) Intergenic regions Synteny Comparative genomics

Having as reference the Next generation Science Standards (NGSS) (National Research Council, 2013), recent studies have shown that despite the school/academic improvements of genetics and genomics content coverage, all the educational stakeholders, including scientist experts and policy-makers, are key players to contribute with recommendations to enhance students' literacy in genetics and genomics (Dougherty et al., 2011; Lontok et al., 2015).

In this context, the present work focus on the identification of specific concepts, presently absent from the curricular contents, but which integration we believe is important to facilitate the understanding of issues addressed in classes, such as gene therapies and GMOs, leading to an engagement of students as citizens and boosting their motivation (Altschul et al., 1990; Nature, 2019). The proposed core concepts were chosen considering their importance for a clear comprehension of the current required notions and to understand up-to-date genomics issues and get acquainted with user-friendly bioinformatics tools:

- *Open Reading Frame* (ORF) is a hypothetical coding sequence with a start and a stop codon (Nature, 2019). This is an absolutely essential concept to understand how raw genome sequences are assembled and annotated.
- *Basic Local Alignment Search Tool* (BLAST) is based on an algorithm which identifies similarities between the query sequence and the sequences deposited in gene banks (Altschul et al., 1990; NCBI, 2020a). BLAST leads to the comprehension of evolutionary relationships and the identification of genes and gene families (NCBI, 2020a).

- *Intergenic regions* are DNA sequences located between codifying sequences/genes (Kahl, 2015). By recognizing the existence of these regions, a better comprehension of how genes are organized in the chromosomes can be achieved while the understanding of gene regulation is improved.

- *Synteny* refers to the preservation of the blocks of genes on chromosomes across different taxa (Duran et al., 2009). This concept is essential to evaluate if gene clusters are conserved, and therefore derived from an ancestral genomic region (CoGe, 2014). This notion is fundamental to approach comparative genomics studies within the scope of evolutionary biology.

- *Comparative genomics* is the scientific field that studies comparatively genomic regions of different taxa in order to disclose affinities among different organisms (Touchman, 2010). This notion allows to explore evolutionary reasoning which evokes the need to hypothesize the presence of identical gene clusters across different taxa.

In order to integrate these specific concepts in genomics teaching practices, a hands-on activity was explored to prospectively develop students' learning skills in genomics and bioinformatics.

3. A Hands-on Approach to Learn Genomics Core Concepts

Previous studies showed that promoting genomics education through practical activities revealed to have a positive impact on students learning (Knox et al., 2003; McQueen et al., 2012; Mil et al., 2010). In this regard, several hands-on proposals for high school can be found in the literature (Conley et al., 2016; Gibson & Cooper, 2017; Lesnik, 2018; Weigel et al., 2014), most mainly centered on *in silico* approaches (Arnold et al., 2017; Martins, Fonseca, et al., 2018; Newman et al., 2016; Wefer, 2003).

The current study addresses the potential of a hands-on bioinformatics-based activity, proposed by us (Martins, Fonseca, et al., 2018), designated as "*Mining the Genome: using Bioinformatics Tools in the Classroom to Support Student's Discovery of Genes*" to promote the learning of the above-mentioned genomics concepts. It is important to emphasize that this activity was designed to provide teachers, generally uneasy with bioinformatics-based exercises, with a suitable didactic instrument which strongly contributes to improve students' knowledge about genomics' concepts and methods. While performing this research driven activity, it is expected for students to strengthen concepts related with protein synthesis and gene regulation (e.g. genome, genes, codons) and also learn new genomics core concepts currently dismissed from curricular contents. Following detailed guidelines, students are driven to identify genes, disclose their genomic context, and hypothesize about their evolution, using up-to-date research platforms (Martins, Fonseca, et al., 2018).

By accessing a comprehensive gene bank database to obtain the specific DNA sequence such as the National Centre for Biotechnology Information (NCBI) database (National Center for Biotechnology Information, n.d.; NCBI Resource Coordinators, 2018), students can understand that genomic information is freely accessible and realize that NCBI is an open access resource.

After retrieving the DNA sequence of interest, students are challenged to use the NCBI ORFfinder (NCBI, 2020b), which allows them to rapidly identify all possible ORFs of a given DNA sequence. This is an important step to understand how to deconstruct a DNA sequence, and identify all possible ORFs, start and stop codons, and getting a glimpse of the routines downstream of the outputs obtained from automatic sequencer machines to the identification and annotation of putative genes.

With all possible ORFs identified, students are asked to verify which of these ORFs might represent putative genes. In this regard, students are introduced to NCBI BLAST tool (NCBI, 2020a). During this task students realize that not all DNA sequences bracketed by a start and a stop codon are coding sequences and that ORFs can be located in different reading frames and oriented in either directions.

To elucidate students about which of the OFRs are actually coding sequences, and also to provide information about the presence of similar putative genes in other taxonomic groups, a BLAST analysis is carried out.

Using the tool MaGe (Magnifying Genomes) of MicroScope (Microbial Genome Annotation & Analysis Platform), an open-access and user-friendly bioinformatics platform for microbial genomics analysis including comparative genomics (LABGeM - French National Sequencing Center, 2020; Vallenet et al., 2006), students can easily retrieve meaningful data namely the genomic coordinates of specific genes; characterize their flaking regions; access their full sequence; determine the reading frame and the coding strand. More importantly, with this *in silico* exercise students get acquainted with tools to comprehensively compare bacterial genomes belonging to different taxa, and intuitively comprehend fundamental concepts of evolution and phylogenomics such as homology and synteny.

4. Conclusion

Taken altogether, this work highlights the benefits and the adequacy of dedicated bioinformatics activities to foster learning of key concepts in genomics education. In addition, the hands-on bioinformatics exercises proposed challenge the students into a research-driven approach while promoting scientific reasoning and literacy. Research is currently being carried out in order to characterize the impact of the mentioned hands-on activity (Martins, Fonseca, et al., 2018) on students' knowledge, interest and attitudes.

Acknowledgements

The authors are grateful to all participants of this study (teachers, students and schools) and to Leonor Martins for the fruitful comments made on the manuscript. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Abrahams, I., & Millar, R. (2008). Does Practical Work Really Work? A study of the effectiveness of practical work as a teaching and learning method in school science. *International Journal of Science Education*, *30*(14), 1945–1969. https://doi.org/10.1080/09500690701749305

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Arnold, M. L., Holman, D., & Zweifel, S. G. (2017). Using Molecular Biology and Bioinformatics to Investigate the Prevalence of Mislabeled Fish Samples. *The American Biology Teacher*, *79*(9), 763–768. https://doi.org/10.1525/abt.2017.79.9.763

CoGe. (2014). *Synteny - CoGepedia*. CoGepedia. https://genomevolution.org/wiki/index.php/Synteny

Conley, J. E., Meisel, A. J., & Smith, J. J. (2016). Using M&M's to Model Sanger's Dideoxy DNA Sequencing Method. *The American Biology Teacher*, *78*(6), 516–522. https://doi.org/10.1525/abt.2016.78.6.516

Črne-Hladnik, H., Peklaj, C., Javornik, B., ˇ Rne-Hladnik, H. C., Košmelj, K., & Hladnik, A. (2009). Assessment of Slovene secondary school students' attitudes to biotechnology in terms of usefulness, moral acceptability and risk perception. *Public*, *18*(6), 747–758. https://doi.org/10.1177/0963662509336761

Dawson, V. M. (2007). An Exploration of High School (12-17 Year Old) Students' Understandings of, and Attitudes Towards Biotechnology Processes. *Research in Science Education, 37*(1), 59-73. https://doi.org/10.1007/s11165-006-9016-7

Ditty, J. L., Kvaal, C. A., Goodner, B., Freyermuth, S. K., Bailey, C., Britton, R. A., Gordon, S. G., Heinhorst, S., Reed, K., Xu, Z., Sanders-Lorenz, E. R., Axen, S., Kim, E., Johns, M., Scott, K., & Kerfeld, C. A. (2010). Incorporating Genomics and Bioinformatics across the Life Sciences Curriculum. *PLoS Biology*, *8*(8), e1000448. https://doi.org/10.1371/journal.pbio.1000448

Dougherty, M. J., Pleasants, C., Solow, L., Wong, A., & Zhang, H. (2011). A comprehensive analysis of high school genetics standards: Are states keeping pace with modern genetics? *CBE Life Sciences Education*, *10*(3), 318–327. https://doi.org/10.1187/cbe.10-09-0122

Duran, C., Edwards, D., & Batley, J. (2009). Genetic maps and the use of synteny. *Methods in Molecular Biology*, *513*, 41–55. https://doi.org/10.1007/978-1-59745-427-8_3

Eijck, M. van. (2010). Addressing the dynamics of science in curricular reform for scientific literacy: the case of genomics. *International Journal of Science Education*, *32*(18), 2429–2449. https://doi.org/10.1080/09500690903473399

EMBL-EBI. (2018). *Train online - What is genomics?* Genomics: An Introduction to EMBL-EBI Resources. https://www.ebi.ac.uk/training/online/course/genomics-introduction-ebi-resources/what-genomics

Gibson, J. P., & Cooper, J. T. (2017). Botanical Phylo-Cards: A Tree-Thinking Game to Teach Plant Evolution. *The American Biology Teacher*, *79*(3), 241–244. https://doi.org/10.1525/abt.2017.79.3.241

Hurle, B., Citrin, T., Jenkins, J. F., Kaphingst, K. A., Lamb, N., Roseman, J. E., & Bonham, V. L. (2013). What does it mean to be genomically literate?: National Human Genome Research Institute Meeting Report. *Genetics in Medicine*, *15*(8), 658–663. https://doi.org/10.1038/gim.2013.14

Kahl, G. (2015). The Dictionary of Genomics, Transcriptomics and Proteomics. In *The Dictionary of Genomics, Transcriptomics and Proteomics*. Wiley-VCH Verlag GmbH & Co. KGaA. https://doi.org/10.1002/9783527678679

Kibirige, I., Rebecca, M. M., & Mavhunga, F. (2014). Effect of practical work on grade 10 learners' performance in science in mankweng circuit, South Africa. *Mediterranean Journal of Social Sciences*, *5*(23), 1568–1577. https://doi.org/10.5901/mjss.2014.v5n23p1568

Knox, K. L., Moynihan, J. A., & Markowitz, D. G. (2003). Evaluation of Short-Term Impact of a High School Summer Science Program on Students' Perceived Knowledge and Skills. *Journal of Science Education and Technology*, *12*(4), 471–478. https://doi.org/10.1023/b:jost.0000006306.97336.c5

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). XThe next-generation sequencing revolution and its impact on genomics. *Cell*, *155*(1), 27. https://doi.org/10.1016/j.cell.2013.09.006

Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*,

12(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

LABGeM - French National Sequencing Center. (2020). *MicroScope Home - MaGe: Microbial Genome Annotation & Analysis Platform - MicroScope - Web Interface System & Specialized Databases for (re)Annotation and Analysis of Microbial Genomes*. https://mage.genoscope.cns.fr/microscope/home/index.php

Lesnik, J. J. (2018). Modeling Genetic Complexity in the Classroom. *The American Biology Teacher*, *80*(2), 140–142. https://doi.org/10.1525/abt.2018.80.2.140

Lewitter, F., & Bourne, P. E. (2011). Teaching Bioinformatics at the Secondary School Level. *PLoS Computational Biology*, *7*(10), e1002242. https://doi.org/10.1371/journal.pcbi.1002242

Lontok, K. S., Zhang, H., & Dougherty, M. J. (2015). Assessing the Genetics Content in the Next Generation Science Standards. *PLoS One*, *10*(7), e0132742. https://doi.org/10.1371/journal.pone.0132742

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annual Review og Genomics and Human Genetics,* *9*(1), 387-402. https://doi.org/10.1146/annurev.genom.9.081307.164359

Marques, I., Almeida, P., Alves, R., Dias, M. J., Godinho, A., & Pereira-Leal, J. B. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r1_id=

Martins, A., Lencastre, L., & Tavares, F. (2018). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. In *3rd International Conference on Teacher Education (INCTE)*. Instituto Politécnico de Bragança. http://hdl.handle.net/10198/17381

Massarani, L., & De Castro Moreira, I. (2005). Attitudes towards genetics: A case study among Brazilian high school students. *Public Understanding of Science*, *14*(2), 201–212. https://doi.org/10.1177/0963662505050992

McQueen, J., Wright, J. J., & Fox, J. A. (2012). Design and Implementation of a Genomics

Field Trip Program Aimed at Secondary School Students. *PLoS Computational Biology*, *8*(8), e1002636. https://doi.org/10.1371/journal.pcbi.1002636

Mendes, A., Rebelo, D., & Pinheiro, E. (2003). *Programa de Biologia e Geologia 11º ou 12ºano(s)*. Ministério da Educação: Departamento do Ensino Secundário.

Mil, M., Boerwinkel, D. J., Buizer-Voskamp, J. E., Speksnijder, A., & Waarlo, A. J. (2010). Genomics education in practice: Evaluation of a mobile lab design. *Biochemistry and Molecular Biology Education*, *38*(4), 224–229. https://doi.org/10.1002/bmb.20397

Munn, M., Skinner, P. O. N., Conn, L., Horsma, H. G., & Gregory, P. (1999). The involvement of genome researchers in high school science education. *Genome Research*, *9*(7), 597–607). https://doi.org/10.1101/gr.9.7.597

National Center for Biotechnology Information. (n.d.). *NCBI*. Retrieved January 22, 2020, from https://www.ncbi.nlm.nih.gov/

National Research Council (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

Nature. (2019). *Open reading frames*. Latest Research and News - Nature. https://www.nature.com/subjects/open-reading-frames

NCBI. (2020a). *BLAST: Basic Local Alignment Search Tool*. https://blast.ncbi.nlm.nih.gov/Blast.cgi

NCBI. (2020b). *Home - ORFfinder - NCBI*. https://www.ncbi.nlm.nih.gov/orffinder/

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Newman, L., Duffus, A. L. J., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

Sadler, T. D., & Zeidler, D. L. (2004). The Morality of Socioscientific Issues: Construal and Resolution of Genetic Engineering Dilemmas. *Science Education*, *88*(1), 4–27. https://doi.org/10.1002/sce.10101

Schuster, S. C. (2008). Next-generation sequencing transforms today's biology. *Nature Methods*, *5*(1), 16-18. https://doi.org/10.1038/NMETH1156

Touchman, J. (2010). Comparative Genomics. *Nature Education Knowledge*, *3*(10), 13. https://www.nature.com/scitable/knowledge/library/comparative-genomics-13239404/

Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., & Médigue, C. (2006). MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Research*, *34*(1), 53–65. https://doi.org/10.1093/nar/gkj406

Wefer, S. H. (2003). Name That Gene: An Authentic Classroom Activity Incorporating Bioinformatics. *The American Biology Teacher*, *65*(8), 610–613. https://doi.org/10.2307/4451571

Wefer, S. H., & Sheppard, K. (2008). Bioinformatics in high school biology curricula: A study of state science standards. *CBE Life Sciences Education*, *7*(1), 155–162. https://doi.org/10.1187/cbe.07-05-0026

Weigel, E. G., DeNieu, M., & Gall, A. J. (2014). Oh, Behave! Behavior as an Interaction between Genes & the Environment. *The American Biology Teacher*, *76*(7), 460–465. https://doi.org/10.1525/abt.2014.76.7.8

Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes

Abstract. *Adapting research-driven routines to the classroom context can promote innovative and motivational learning environments. Using a case-study approach, we propose a set of bioinformatics-based activities supported by a tutorial video aiming to identify genes and disclosing their genomic context in different species. The rationale is to strengthen teachers' competencies to introduce bioinformatics resources and tools (e.g. NCBI, ORFfinder, BLAST and MaGe) in their teaching practices. By doing so, teachers will ultimately enhance students' understanding of how genomic data mining and comparative genomics are instrumental for biological research.*

Keywords. Bioinformatics, Comparative Genomics, Evolution, High School, Motivation

1. Introduction

Nowadays computers have a central function in scientists' daily routine. A personal computer connected to the web is all that it takes to access a myriad of bioinformatics resources capable of deconstructing genomic information into biologically meaningful data. Bioinformatics provides tools to comprehensively analyze and save large amounts of biological data that would be impossible to investigate without informatics-based approaches (Bloom, 2001; Madigan et al., 2018). Here, we present a series of bioinformatics activities that enable students, under the guidance of their teachers, to query an unknown DNA sequence, mimicking a real research scenario. Activities that encourage research-driven problems appear to be a stimulus to students' interest in scientific careers (STEM), since research-inspired activities allow them to get familiar with scientific professions and the academic training required to pursue them (Kovarik et al., 2013).

In order to reconcile simple and yet curriculum-oriented bioinformatics activities intended for high school students (15-17 years old) with high learning impact and didactic value, an inquiry-based scenario structured in four bioinformatics exercises was designed. Besides having a positive impact on students' engagement and motivation (Campbell, 2003), the educational value of these activities extends to the multiple curricular exploration opportunities they offer. For instance, simply by selecting the query

DNA sequences to be used, teachers can address a plethora of topics framed in the *Next Generation Science Standards*, such as gene regulation, evolution and drug-resistance (Brock, 1998; Cooper, 2015; National Research Council, 2013; Moss, 1997; Newman et al., 2016; Taylor et al., 2014).

2. Online Resources

The bioinformatics applications used in the exercises detailed below are open access and web-based, with user-friendly interfaces that run in common web browsers of PC and Mac computers. Although the applications chosen are hosted in long-established web-based platforms that are widely used and currently indispensable in daily research routines, it is important to instruct students about the evolving dynamics of these bioinformatics applications, resulting from the addition of more data, the development of new resources, or the display of increasingly intuitive interfaces.

A pilot trial of these bioinformatics activities was carried out in a classroom setting with the collaboration of fourteen teachers from six schools and involving a total of 387 high school students (15-18 years old).

3. Learning Objectives

Specific learning objectives are detailed after each exercise. Through all these activities, students

- Strengthen their knowledge of concepts such as genome, chromosomes, genes (structural, operator, repressor, regulator, promoter), start and stop codons, and operons;

- Learn new concepts such as open reading frames, synteny, and comparative genomics; and

- Improve their computational skills and increase their digital literacy.

4. Class Workflow

To adapt the bioinformatics activities to a classroom context properly integrated in the high school curricula, the exercises were designed in collaboration with the teachers who took part in the pilot trial. Taking into account teachers' suggestions we propose a class workflow comprising four parts (I-IV), as schematically represented in Figure II - 1 and detailed below. To further assist teachers in implementing the class workflow a tutorial video detailing the four parts was produced (see supplemental file). The estimated times correspond to the average time required by teachers to implement the full set of activities described fellow with their students. Regardless of the suggested timeline, it is important

to emphasize that each teacher may easily reschedule the class workflow according their teaching agenda either by cutting one or more of the four parts, or, alternatively, by stimulating the students' discussion after each exercise.

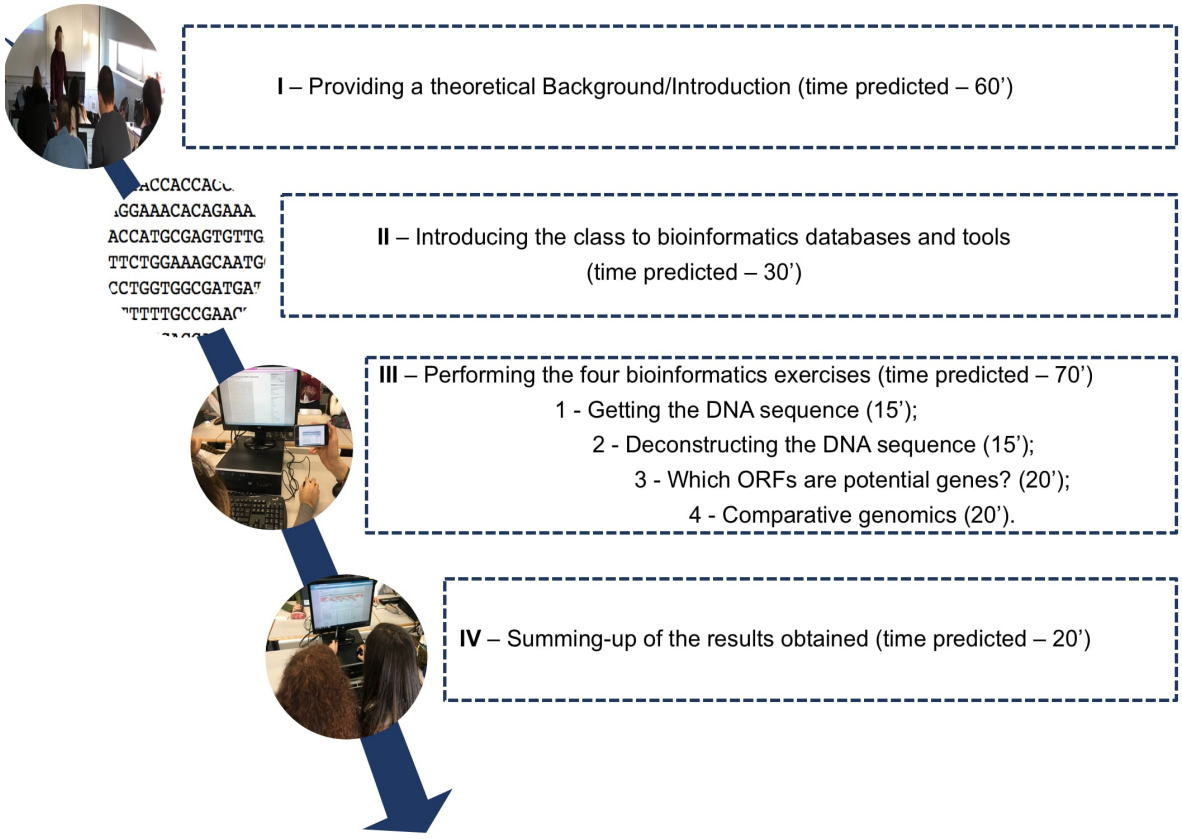


Figure II-1. Proposed class workflow and timeline, taking into account the feedback of 14 in-service teachers who implemented the exercises in their high school classes as a pilot trial.

I – Setting up the theoretical background (estimated time – 60 minutes): The teacher emphasizes the importance of identifying genes from a genomic sequence. Besides recalling basic concepts such as genome, chromosomes, genes (structural, operator, repressor, regulator, promotor), and operons, students are introduced to important notions, namely start and stop codons, open reading frames (ORFs), synteny and comparative genomics.

II – Introduction to bioinformatics databases and tools (estimated time – 30 minutes): The teacher highlights the importance of bioinformatics by explaining the exercises and introducing students to the bioinformatics resources and tools they will use, namely NCBI database, NCBI ORFfinder, NCBI BLAST and Microscope (MaGe). A tutorial video (see supplemental file) should help teachers in this task and assist students throughout the exercises.

III – Bioinformatics exercises (estimated time – 70 minutes): Students carry out the exercises autonomously with the teacher's supervision to identify difficulties and answer questions.

IV – Discussion of the results (estimated time – 20 minutes): The class discusses the results obtained in each exercise and assay to draw conclusions. Ultimately, the teacher might challenge the students to explore other case studies and study different genomic regions. In addition, we should not neglect students endeavor to explore autonomously the bioinformatics resources, particularly taking into account their user-friendly and intuitive interfaces. In fact, during the pilot trial, we observed that some students took the initiative to extend their *in silico* experiments beyond the assigned activities by pursuing their own research queries, as for instance: "What is the size of the genome of a spider?"; "Are virus genomes such as HIV also available at this database?"; "Let's search for the gene coding for insulin."

5. Bioinformatics Exercises

The bioinformatics-based activities described below are structured according to four distinct exercises (see supplemental file): 1 - getting the target DNA sequence; 2 – looking for ORFs; 3 - deciding which of the retrieved ORFs are likely to be genes; and 4 - analyzing the gene(s) identified within their expected genomic context. Having in mind that laboratory-based activities should meet the curricular agenda and acknowledging the fact that *lac* operon is a common example for teaching gene regulation, the query DNA sequence chosen to exemplify these exercises corresponds to *lacI* and flanking regions. Furthermore, to frame the bioinformatics-based activities in an inquiry-based approach, all exercises start with a guiding question.

1 - Getting the DNA Sequence

This initial exercise aims to answer the question: *"How does one access a comprehensive gene bank database to obtain the specific DNA sequence to be studied?"*

1.1. Access NCBI website: http://www.ncbi.nlm.nih.gov/.

1.2. Choose Genome in menu next to the search box.

1.3. Search by "*E.coli*".

1.4. At the beginning of the new page, select Reference Genome by clicking the *E.coli* strain K12.

1.5. Scroll down and click on the accession number corresponding to *E.coli* strain K12 in the Reference Sequence command to retrieve the full genome.

1.6. Choose the FASTA format.

1.7. Open the selection box Change region shown and type down the coordinates 366001-368041.

1.8. Copy, paste and save the sequence in a Word or Notepad document.

Learning objectives. Through the exploration of the comprehensive bioinformatics database NCBI, students learn

- how the database is organized, its complexity, and

- how to search for DNA sequences and gene sequences for different organisms.

2 - Deconstructing the DNA Sequence

This exercise was planned to instruct students how to go from an unknown DNA sequence to the identification of hypothetical coding sequences. Students are introduced to the notion of ORFs, which frequently escapes the scientific lexicon of elementary and high school biology curricula, but which is instrumental for answering the question: *"How is a new DNA sequence deconstructed?"*

2.1. Access NCBI ORFfinder: http://www.ncbi.nlm.nih.gov/orffinder/.

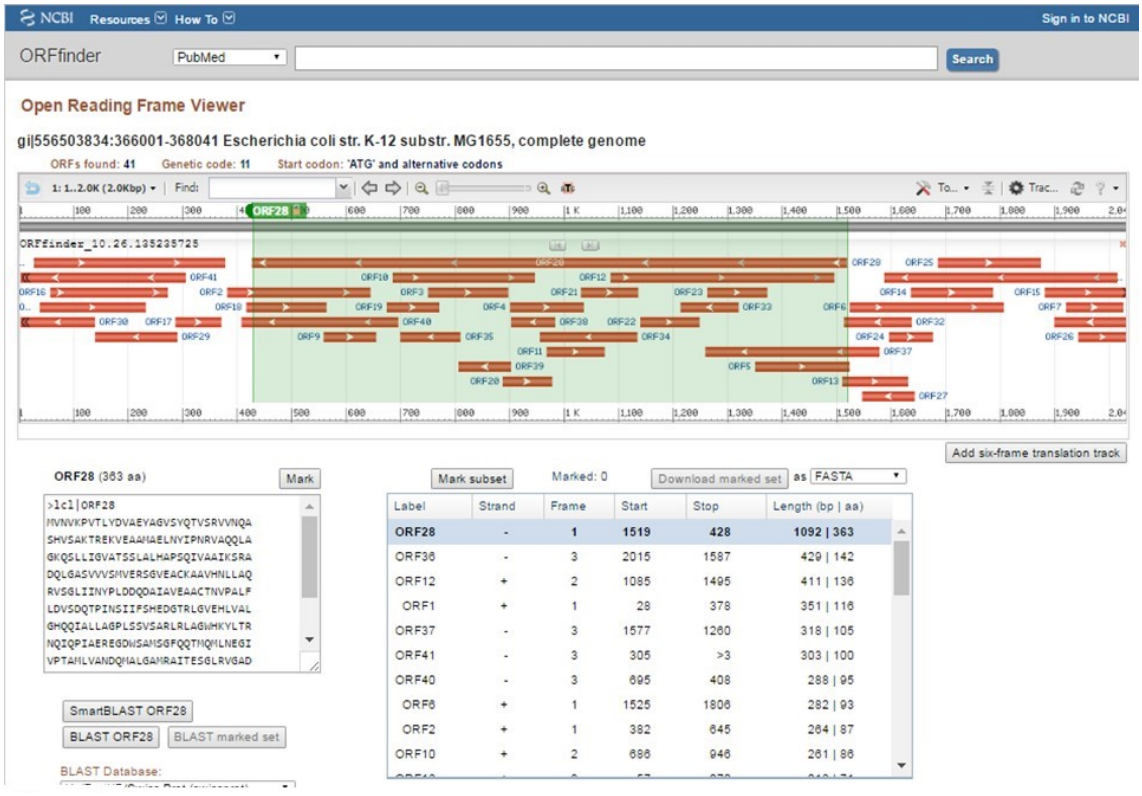
2.2. Paste the sequence previously saved as Word or Notepad document into the text box provided.

2.3. Choose the genetic code: 11. Bacterial, Archaeal and Plant Plastid.

2.4. Choose the option: "ATG" and alternative initiation codons.

2.5. Click Submit.

2.6. Analyze the obtained results (Figure II - 2).



https://www.ncbi.nlm.nih.gov/orffinder/

Figure II-2. ORFfinder output at NCBI, disclosing all possible Open Reading Frames (ORFs) and their direction within the query DNA sequence. In addition to the graphic view, details such as ORF coordinates, length, strand and frame are highlighted in the table below. By selecting each ORF, it is possible to obtain the translated aminoacid sequence. Immediate BLAST of each ORF may be executed with the command BLAST ORF.

Learning objectives. With this exercise, students

- recognize the six different reading frames in a DNA sequence,

- understand the meaning of ORF, and

- recognize the importance of start and stop codons for identifying all possible ORFs.

3 - Which ORFs Are Potential Genes?

Basic Local Alignment Search Tool (BLAST) is a powerful algorithm capable of finding similarities between a query sequence (DNA or a protein sequence) and the sequences available in databases (Altschul et al., 1990). Using this application, the students can address the following questions: *"Which of the ORFs retrieved in the previous exercise are probable genes? Which ORFs are unlikely functional coding sequences?"*

3.1. Select one ORF to study (example: ORF 28).

3.2. Start BLAST of the selected ORF by clicking on BLAST ORF.

3.3. Click on BLAST in the new page opened.

3.4. Identify the gene (Figure II - 3).

3.5. Repeat the procedure for other ORF's and analyze the results obtained.

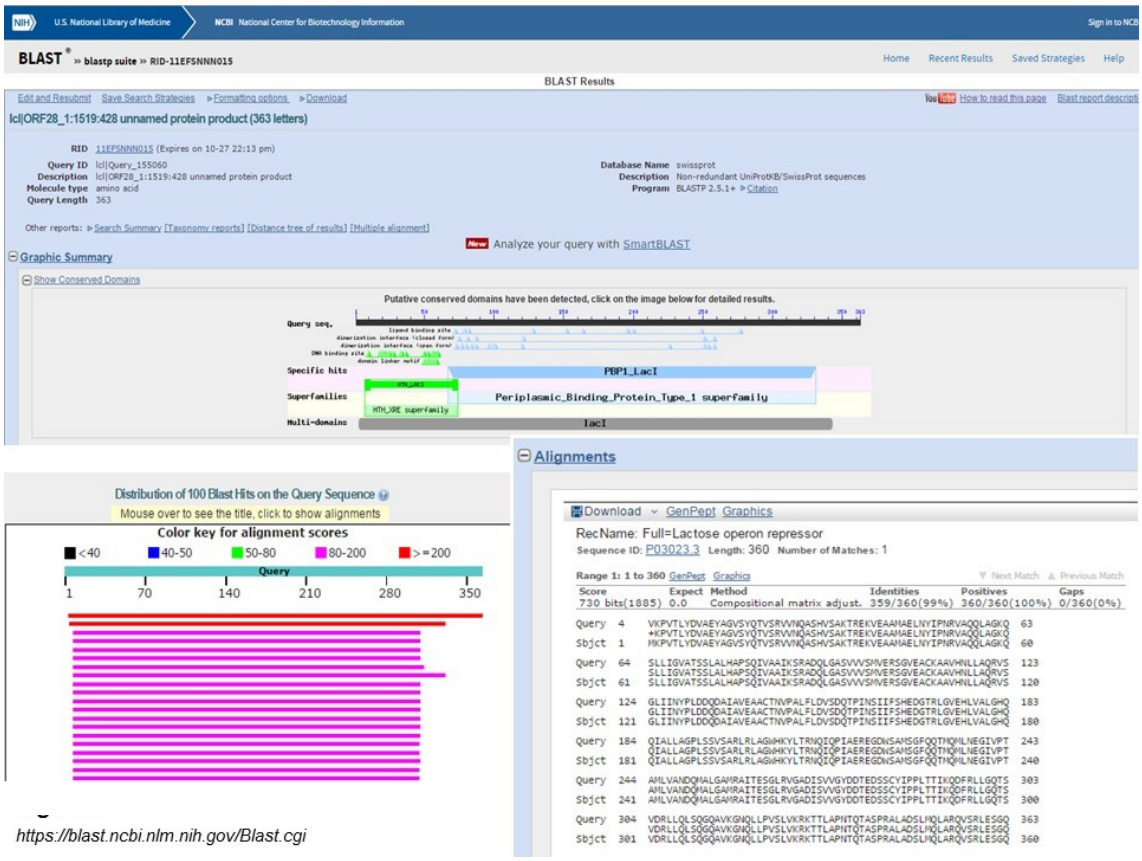


Figure II-3. BLAST output at NCBI, highlighting the similarity scores between the query ORF and the 100 best BLAST hits retrieved in the database. Clicking over a line displays the alignment between the query sequence and the subject sequence allowing identification of differences between the two sequences.

Learning objectives. Students learn that

- not all DNA sequences bracketed by a start and a stop codon (i.e. ORFs) are coding sequences,

- ORFs can be located in different reading frames and oriented in either direction, and

- scrutinizing gene banks by a BLAST search is an effective approach for identifying putative genes among retrieved ORFs.

Students can discuss possible scenarios to explain a BLAST search in which no similarities are found.

4 - Comparative Genomics

To fully exploit the potential of this activity, the fourth exercise compares the presence of the identified gene or genes, their genomic context, and their occurrence across different taxa. Using MaGe, a robust comparative genomics platform (Vallenet et al., 2006), the students further confirm the identity and putative function of the gene(s)

determined during the BLAST search. Students might ask, *"Is there any evolutionary relationship to explain the occurrence of the studied genes across different taxa?"*

4.1. Access *MicroScope* website: https://www.genoscope.cns.fr/agc/microscope/home/index.php.

4.2. Choose *Escherichia coli* K12 and select Load into genome browser.

4.3. To identify the gene, search for "*lacI*" and click: Move to.

4.4. Identify *lacI* gene putting the mouse over each red bar.

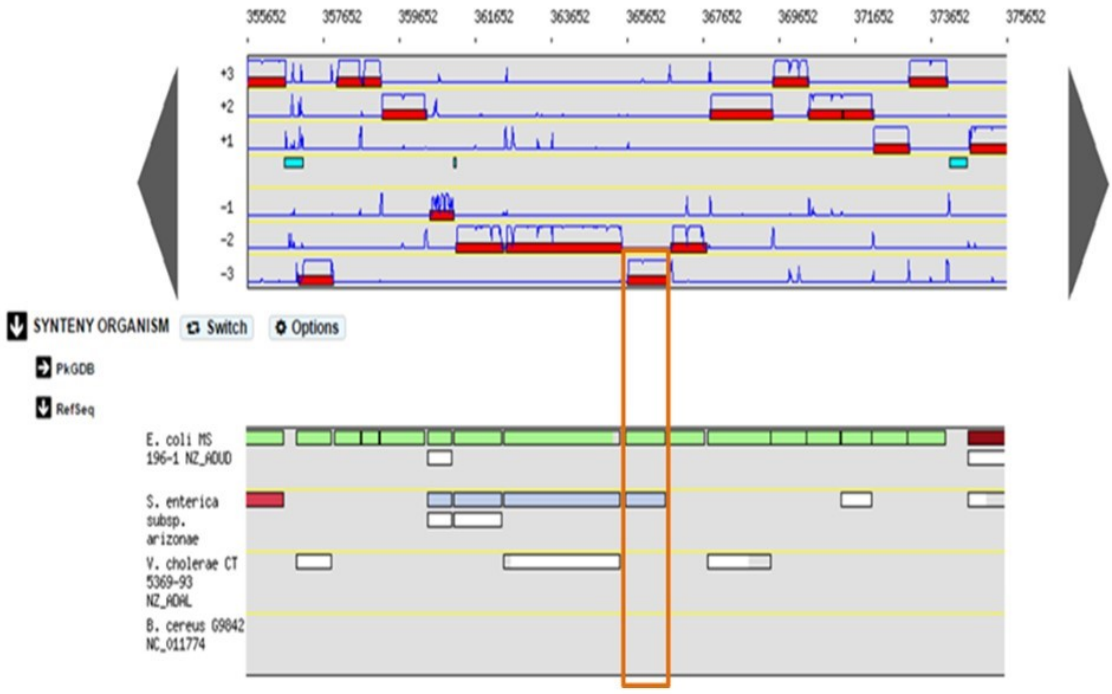
4.5. Select options menu.

4.6. In the new window opened, look for the section Viewer Comparative Map default and choose synteny.

4.7. In the section PkGDB Organism Synteny, press the button CTRL and choose *Bacillus anthracis*, another *Escherichia* species, *Salmonella bongori*, *Shigella sonnei* and *Vibrio cholera*.

4.8. Click Save options.

4.9. Compare the presence and the function of the gene in different taxa (Figure II - 4).



http://www.genoscope.cns.fr/agc/microscope/home/index.php

Figure II-4. Comparative genomics analysis carried out using MaGe. The genes and corresponding reading frames (+3, +2, +1, -1, -2, -3) of the query genes are shown at the top. Below is an outline of other bacteria with which the query gene(s) are being compared.

Learning objectives. Through this simple comparative genomics analysis, students learn

- to localize their target gene(s) within the chromosome,

- to identify the genomic features of the flanking regions,

- to determine gene homologies with selected taxa, and

- concepts such as synteny, homology, insertions, deletions, and horizontal gene transfer.

Additional Remarks

The pilot trial showed that internet access was not a limitation when implementing these activities at schools. Nevertheless, teachers could easily choose to exclude one of the exercises or, alternatively, to challenge the students to carry them out as homework, and later resume the bioinformatics exercises in the classroom.

Acknowledgements

Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015). The authors are grateful to all the participant schools and school teachers for the opportunity to implement the bioinformatics exercises detailed in this work, which contributed to improving the described activity.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bloom, M. (2001). Biology *in silico*: The Bioinformatics Revolution. *The American Biology Teacher*, *63*(6), 397–403. https://doi.org/10.2307/4451145

Brock, D. L. (1998). Now You See It, Now You Don't!: Making Regulation of Gene Expression Come Alive for All Students. *The American Biology Teacher*, *60*(4), 288–290. https://doi.org/10.2307/4450474

Campbell, A. M. (2003). Public access for teaching genomics, proteomics, and bioinformatics. *Cell Biology Education*, *2*(2), 98–111. https://doi.org/10.1187/cbe.03-02-0007

Cooper, R. A. (2015). Teaching the Big Ideas of Biology with Operon Models. *The American Biology Teacher*, *77*(1), 30–39. https://doi.org/10.1525/abt.2015.77.1.5

Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Madigan, M. T., Bender, K. S., Buckley, D. H. (Daniel H., Sattley, W. M., & Stahl, D. A. (2018). *Brock biology of microorganisms* (15th ed.). Pearson Education/Benjamin Cummings.

Moss, R. (1997). A Discovery Lab for Studying Gene Regulation. *The American Biology Teacher*, *59*(8), 522–526. https://doi.org/10.2307/4450370

National Research Council (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

Newman, L., Duffus, A. L. J., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

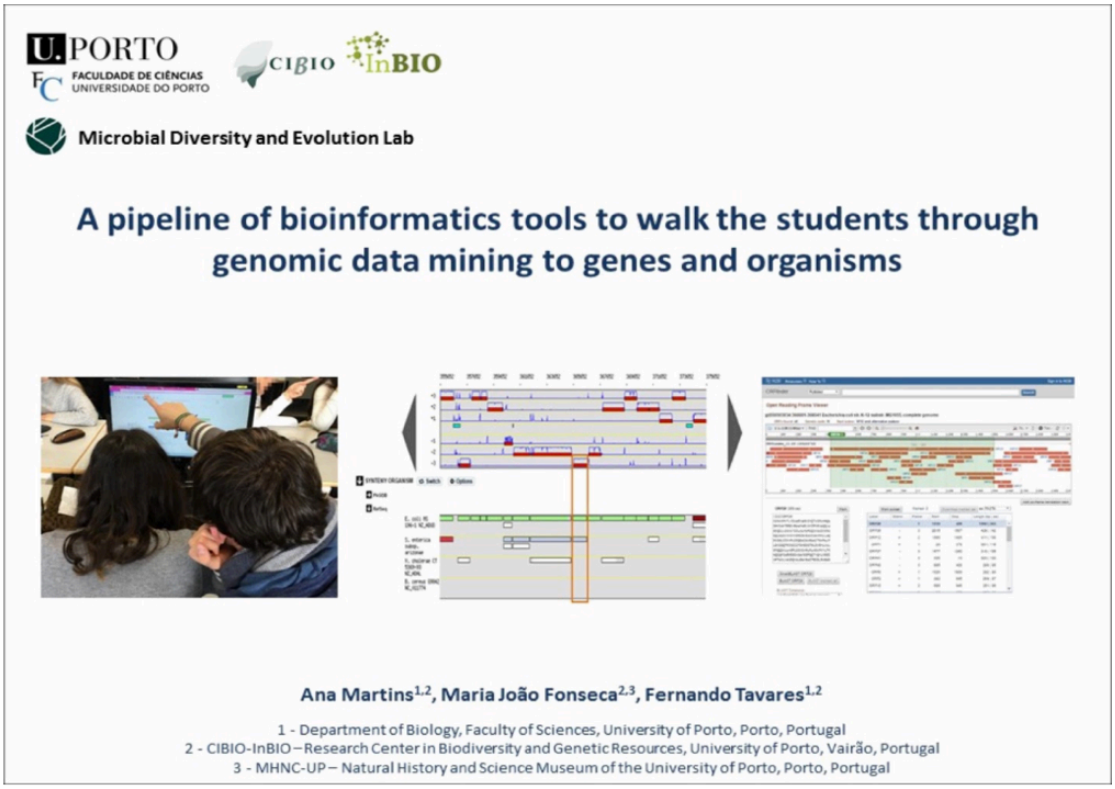
Taylor, J. M., Davidson, R. M., & Strong, M. (2014). Drug-resistant tuberculosis: A genetic analysis using online bioinformatics tools. *American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Vallenet, D., Labarre, L., Rouy, Z., Barbe, V., Bocs, S., Cruveiller, S., Lajus, A., Pascal, G., Scarpelli, C., & Médigue, C. (2006). MaGe: a microbial genome annotation system supported by synteny results. *Nucleic Acids Research*, *34*(1), 53–65. https://doi.org/10.1093/nar/gkj406

Supplemental file: Tutorial video to support the implementation of bioinformatics-based activities proposed in the current works.

Tutorial video available at:

https://drive.google.com/file/d/1WrtjzLHzKI7nLALtVnmqy6WhPnkUsQSR/view



Chapter III

Students' Literacy, Interest and Attitudes towards Bioinformatics in Formal and Non-formal Educational Contexts

Chapter III includes the following publications:

Martins, A., Fonseca, M. J., Lemos, M., Lencastre, L., & Tavares, F. (2020). Bioinformatics-Based Activities in High School: Fostering Students' Literacy, Interest and Attitudes on Gene Regulation, Genomics and Evolution. *Frontiers in Microbiology, 11,* 578099. https://doi.org/10.3389/fmicb.2020.578099

Martins, A., Lencastre, L., & Tavares, F. (2018). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education*. (pp. 309–317). Hands-on Science Network. ISBN: 978-84-8158-779-1

Part of the results from this chapter were presented at:

Martins, A., Lencastre, L., & Tavares, F. (2019). Bioinformatics in Secondary Education: From Wishful Thinking to Reality. Microbiotec 19. Coimbra, 5th-7th December 2019., 9th – 11th July 2018.

Bioinformatics-Based Activities in High School: Fostering Students' Literacy, Interest and Attitudes on Gene Regulation, Genomics and Evolution

Abstract. *The key role of bioinformatics in explaining biological phenomena calls for the need to rethink didactic approaches at high school aligned with a new scientific reality. Despite the several initiatives to introduce bioinformatics in the classroom, there is still a lack of knowledge on their impact on students' learning gains, engagement and motivation. In this study, we detail the effects of four bioinformatics laboratories tailored for high school biology classes named "Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes" on literacy, interest and attitudes on 387 high school students. By exploring these laboratories, students get acquainted with bioinformatics and acknowledge that numerous bioinformatics tools can be intuitive for beginners. Furthermore, introducing comparative genomics in their learning practices contributed for a better understanding of curricular contents regarding the identification of genes, their regulation and how to make evolutionary assumptions. Following the intervention students were able to pinpoint bioinformatics tools required to identify genes in a genomics sequence and, most importantly, they were able to solve genomics-related misconceptions. Overall, students revealed a positive attitude regarding the integration of bioinformatics-based approaches in their learning practices, reinforcing their added value in educational approaches.*

Keywords. Bioinformatics, Comparative Genomics, Gene Regulation, High School, Genomic Literacy

1. Introduction

Bioinformatics, understood as the use of computational resources to categorize massive raw data and retrieve meaningful information from datasets, has gain a primordial utility in scientists' daily routine (Sadek, 2004). This paradigm of biological research cannot be disregarded when seeking to promote a scientifically informed society. Indeed, it demands the improvement of curricular and educational resources at middle and high school educational levels, based on initiatives validated by focused science education research.

Learning by accessing online bioinformatics resources in the classroom has already proved to have a beneficial impact on students' ability to build up and mobilize scientific contents,

namely related to drug-resistance, phylogenetic trees or genetic expression (Amenkhienan & Smith, 2006; Machluf et al., 2017; Newman et al., 2016; Taylor et al., 2014). In addition, the introduction of bioinformatics at high school enhances the learning of new information, through novel technologies and recruits resources used in research laboratories, serving as a stimulus to spark students' future interest in scientific careers (Kovarik et al., 2013; Machluf et al., 2017).

Despite the various initiatives across Europe to support teachers and students to integrate bioinformatics-based approaches in their classes, these remain sporadic and are still not implemented consistently. Recent studies have called attention to the importance of a joint effort by all stakeholders (e.g. research institutions, governmental entities, teachers, trainers and researchers) to deliver an action plan that can lead to bioinformatics dissemination in schools in a wider, more structured and cohesive manner (Attwood et al., 2017; Campbell & Nehm, 2013; Koch & Fuellen, 2008). Recent reports call for more educational assessments to strengthen the positive impact of bioinformatics-based activities on students' scientific and digital literacy, providing a rationale to incorporate bioinformatics in the curriculum (Campbell & Nehm, 2013; Dudley & Butte, 2009; Machluf et al., 2017; Machluf & Yarden, 2013; Magana et al., 2014; Marques et al., 2014)

This study aims to address the educational impact on high school students of a set of activities developed to introduce basic bioinformatics analysis used to deconstruct a bacterial genomic sequence into its coding genes (Martins, Fonseca, et al., 2018), using purposely tailored evaluation instruments. The main research question driving this investigation was: *are there significant changes in high school students' scientific and digital literacy, interest and attitudes towards gene regulation, genomics and evolution after performing bioinformatics-based activities?*

2. Materials and Methods

2.1. Participants

The sample studied included a group of 387 students and 11 teachers from five public and private schools in Porto and Lisboa, Portugal. Fourteen 11th grade biology and geology classes (students' age: 16-17 years old) and five 12th grade biology classes (students' age: 17-18 years old), comprising 167 male and 220 female students, were involved in this study. Students' average age was 16.34 ± 0.67 years. The study included an experimental group (n=292) with 123 male students and 169 female students (average age: 16.27 ± 0.68 years) from 14 classes, and a control group (n=95) including 44 male students and 51 female students (average age: 16.54 ± 0.62 years) from five classes.

Students participated in the project as part of their science classes and taking into account all ethical requirements, the project was institutionally approved by each school's Directive Board. Upon entering the project, the participants were invited to take part in the study and informed of its nature and aims, being assured that all the data collected were to be processed and analyzed anonymously. Students were given the chance to participate in the project without participating in this specific study.

2.2. Didactic Instrument: Bioinformatics Laboratories

A set of bioinformatics-based activities previously proposed by Martins, Fonseca, et al. (2018) to identify genes from a bacterial genomic sequence and disclose their genomic context in different species was chosen as the didactic instrument. A tutorial video (https://drive.google.com/file/u/1/d/1WrtjzLHzKI7nLALtVnmqy6WhPnkUsQSR/view) provides teachers and students with a detailed road map of the sequential bioinformatics resources needed to deconstruct a 2kb genomic region of *Escherichia coli* and determine its occurrence across different bacteria taxa and hypothesize about its evolution. Participants were initially instructed to select a particular *E. coli* strain (*Escherichia coli* str. K-12 substr. MG1655, Accession number: NC_000913.3) and a specific 2kb genomic region, to ensure that all of them would be working with the same genomic sequence, allowing for a more efficient teacher supervision and facilitating subsequent analysis. In fact, the 2kb sequence proposed includes the *lac* operon, which is the paradigm used to introduce gene expression and regulation at the high school. This provides a meaningful curricular framing for these activities and is aligned with students' previous knowledge. Furthermore, it is important to emphasize that implementing bioinformatics exercises framed within the curriculum was a main concern of the participant teachers. Currently the Portuguese biology curricula for the 11th and 12th grades include contents related with DNA and protein synthesis (for example, transcription, translation, and start and stop codons), as well as evolution (Mendes et al., 2003), and genetic expression (Mendes et al., 2004). These topics are also comprised in the Next Generation Science Standards (NGSS) (National Research Council, 2013). While these curricular topics are frequently focused on eukaryotic models, bacterial genomes were chosen as an educational instrument for this study having in mind that bacteria stand for the most represented domain in genome databases, reflecting its high taxonomic diversity, and may be easily recruited by ingenious bioinformatics platforms with graphical and user-friendly interfaces using a windows or mac browser. In addition, bacterial genomes are frequently restricted to a single replicon (i.e. the chromosome), besides having a small-sized haploid genome that favors comparative genomics and contributes to strength students' knowledge on bacteria, fostering their motivation and interest on microbiology related topics, presently poorly explored in high school.

The bioinformatics resources used include the genome database, Open Reading Frames Finder (ORFfinder), and Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI) (National Research Coordinators, 2018; Altschul et al., 1990), and the genome browser of Magnifying Genomes (MaGe) that is part of MicroScope, a comparative genomics platform (Vallenet et al., 2013). Before starting with the *in silico* laboratories the teachers work through basic and already known concepts such as genome, genes, start codons, stop codons, and operons with the class, and introduce new notions, such as open reading frames (ORFs), synteny, and comparative genomics (Figure III - 1) (Martins & Tavares, 2018). This is particularly important since these new notions, presently absent from the curricula, are instrumental to understand the data retrieved by the students when performing the bioinformatics exercises proposed (Martins & Tavares, 2018).

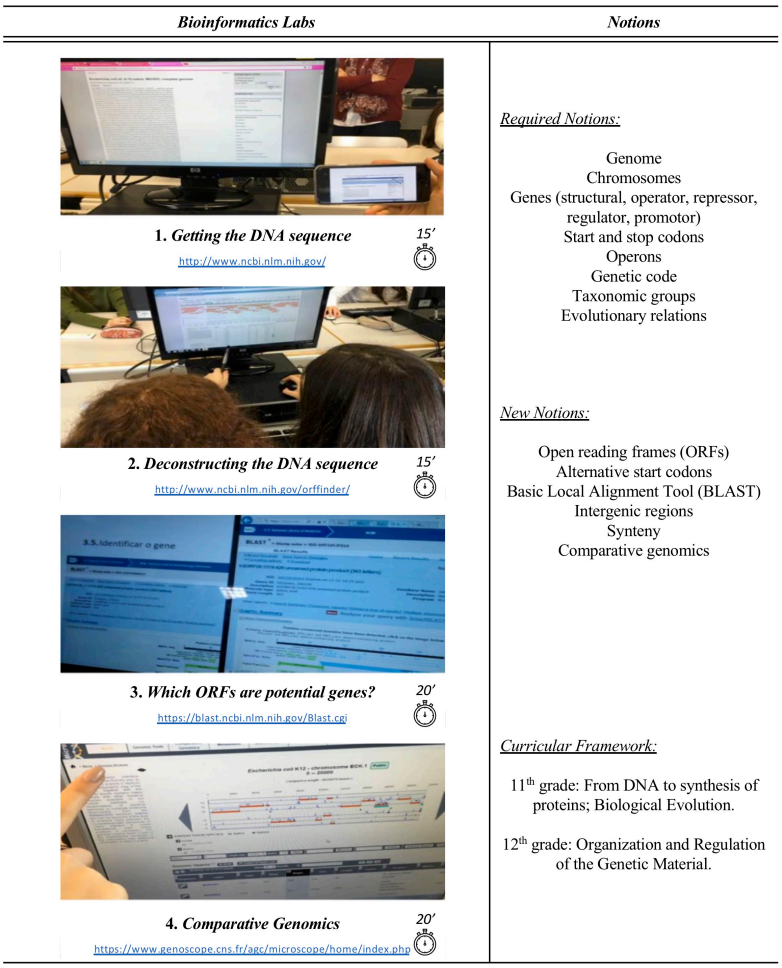


Figure III-1. Bioinformatics laboratories framed within the curricular biology contents for high school to reinforce genomics topics currently required and to introduce new core concepts.

2.3. Research Design and Methodology

To implement bioinformatics-based activities as a successful didactic instrument it is crucial to engage both teachers and students in the selection of the activities to ensure that these are meaningful and adjusted to the curricular contents (Marques et al., 2014). In this regard, the design of the bioinformatics-based activities proposed by Martins, Fonseca, et al. (2018) took into account teachers' contribution in revising and piloting the proposed educational resources with their students (Figure III - 2). To lighten the burden for teachers, a dedicated webpage (https://bioinformaticaaula.wixsite.com/bioinformatica-pt) was developed to provide them with resources that introduced the bioinformatics tools and the new concepts to be addressed.

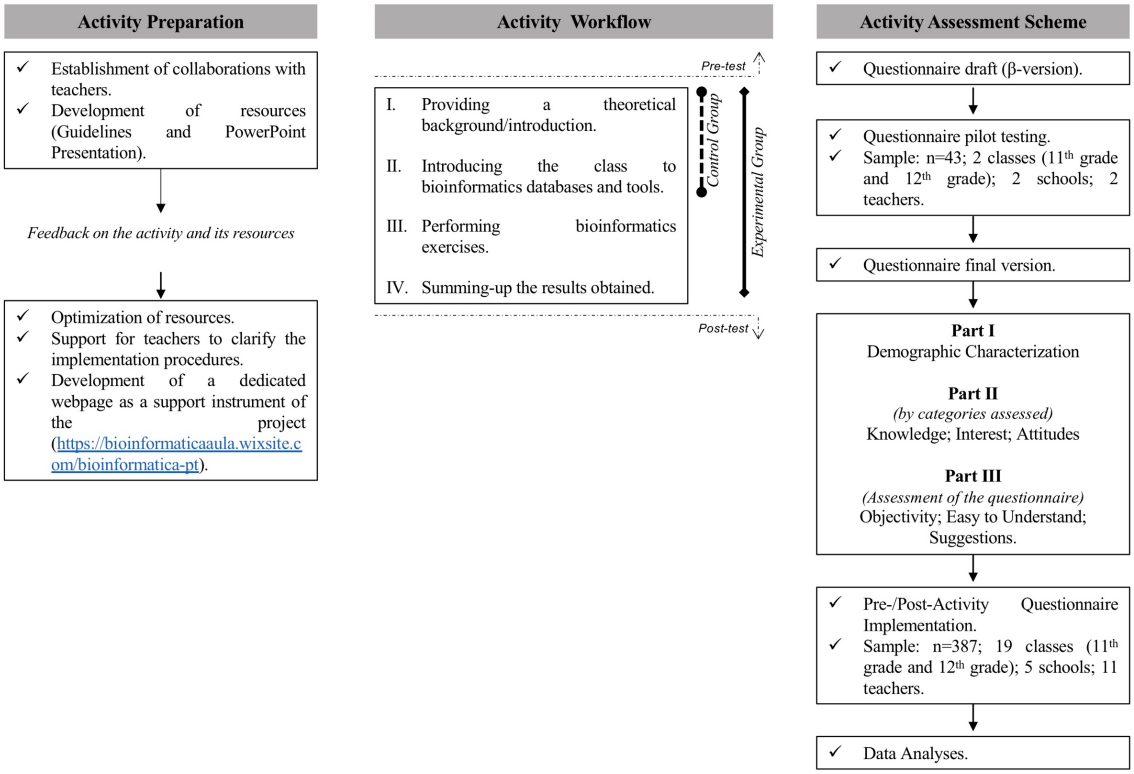


Figure III-2. Experimental design for preparation, implementation and assessment of bioinformatics-based activities.

The workflow of the bioinformatics activities includes four parts (Figure III - 2). Firstly, teachers provide the knowledge background about gene regulation, genomics and evolution. Secondly, students are introduced to the bioinformatics databases and tools to be used (namely NCBI database, NCBI ORFfinder, NCBI BLAST and Microscope (MaGe)). And thirdly, the bioinformatics exercises are performed. These exercises were set to meet the curricular requirements for the topic and, given the novelty of bioinformatics for these students (and teachers), guidelines were prepared to provide a comprehensible workflow to address the research questions outlined. This allowed to prevent students from becoming overwhelmed by

the wide plethora of choices of links and commands available in the platforms mentioned before. In the fourth and final stage of implementation, the results obtained in each exercise were discussed with the students and conclusions were drawn.

During the implementation of the activities, a member of the research team (Martins) was present to identify misconceptions and reasoning difficulties, as well as to check the participants' engagement and interaction, and to carry out qualitative observations useful to improve the robustness of the interpretations made.

A quasi-experimental pre-/post design, with a control and an experimental group, was set up. The control group included classes exclusively exposed to the first two parts of the intervention, i.e. the introductory lectures about the scientific questions and the bioinformatics databases and resources (Figure III – 2 – Workflow I and II). In turn, the experimental group was exposed to the full set of the bioinformatics activities, i.e. from the introductory lectures to the bioinformatics laboratories and the interpretation of the results (Figure III – 2 – Workflow I to IV). To mitigate possible bias effects, the control group classes were from the same schools, from the same education levels and taught by the same teachers as the experimental group classes. The comparison between the performance of students in the control group and the experimental group was intended to test the educational impact of the practical bioinformatics-based activities. In this regard, the control group was taught only through expositive teaching (Figure III – 2 – Workflow I and II) and the experimental group was exposed to the same lectures as the control group plus the practical component (Figure III – 2 – Workflow I to IV).

2.4. The Questionnaire

To assess the educational impact of integrating the mentioned bioinformatics-based activities in high school, a specific and comprehensive questionnaire including open-ended questions, dichotomous questions and Likert-type scales, was designed (Figure III - 3).

The questionnaire was structured according to three main dimensions: knowledge, interest and attitudes. The knowledge-related questions (Q1, Q2, Q4, Q5, Q6, Q7, Q8.5) aimed to characterize students' literacy regarding gene regulation, comparative genomics, bioinformatics and its usefulness for scientific research. Students' interest (Q3 and Q9) was measured by their perception of the role of bioinformatics in tackling different biology research questions and by their awareness about the scientific disciplines addressed in the *in silico* activities, namely genetics, genomics and evolution. Students presently attending high school are part of the so called *iGeneration* (iGen) which is characterized by being highly motivated to use of technology in their daily lives (Quinn & Oldmeadow, 2013; Rosen et al., 2010). Having this in mind, a question (Q8) was added to depict students' attitudes towards the use of

computer/technological devices to study, and to assess their motivation to access bioinformatics tools inside or outside the classroom.

Knowledge

⇒ **Q1:** Have you heard about bioinformatics?
  → **Q1.1:** If so, describe what is bioinformatics for you.
⇒ **Q2:** Imagine the following situation: "*As a researcher, you sequence a genomic fragment. Do you have any idea how you would proceed to identify the gene(s) present?"*
  → **Q2.1:** If so, indicate the main procedures that would follow to identify the gene(s) present in that sequence.
⇒ **Q4:** What is genomics for you?
⇒ **Q5:** Have you heard about comparative genomics?
  → **Q5.1:** If so, define comparative genomics.
⇒ **Q6:** Answer with True/False/Don't Know:
  → **Q6.1:** Databases, known as genebanks, are free access resources.
  → **Q6.2:** All citizens have access to the main genomic databases.
  → **Q6.3:** All bioinformatics tools require programing skills.
  → **Q6.4:** Bioinformatics tools are essential to molecular biology studies.
⇒ **Q7:** Rate your agreement with the following statements (1 – *I totally disagree;* to 5 – *I totally agree*):
  → **Q7.1:** Different taxonomic groups of bacteria have genes in common.
  → **Q7.2:** There are different initiation codons.
  → **Q7.3:** There is a specific genetic code for bacteria.
  → **Q7.4:** All the bacterial genes are known.
⇒ **Q8.5:** In case you have used the computer / technological devices to access bioinformatics resources please indicate which bioinformatics tools you used.

Interest

⇒ **Q3:** How important do you think bioinformatics is to the following activities (1 - *Not important at all*; to 5 - *Very important*):
  → **Q3.1:** To identify genes.
  → **Q3.2:** To store genomic data.
  → **Q3.3:** To study the evolutionary relations between organisms.
⇒ **Q9:** Rate the importance of the following practices (1 – *Not important at all*; to 5 – *Very important*):
  → **Q9.1:** Practical work using digital tools (e.g. virtual labs, videos, use of interactive applications, etc.) in the classroom.
    ◊ **Q9.1.1:** Justify your choice.
  → **Q9.2:** Study of genomes and gene regulation in bacteria.
    ◊ **Q9.2.1:** Justify your choice
  → **Q9.3:** Study of phylogeny / evolution of bacteria.
    ◊ **Q9.3.1:** Justify your choice.
  → **Q9.4:** Using bioinformatics tools in the class.
    ◊ **Q9.4.1:** Justify your choice.

Attitudes

⇒ **Q8:** How often do you … (1 – *Never*; to 5 – *Always*):
  → **Q8.1:** Use the computer / technological devices for autonomous study outside the classroom.
  → **Q8.2:** Use the computer / technology devices in the classroom to study.
  → **Q8.3:** Use the computer / technological devices to access bioinformatics tools outside the classroom.
  → **Q8.4:** Use the computer / technological devices to access bioinformatics tools in the classroom.

Figure III-3. The questionnaire used in this study included demographic characterization of the participants and items to assess students' knowledge, interest and attitudes.

The questionnaire developed was piloted in two high school classes (n=43 students) (Figure III - 2), which, as recommended by several authors (Connelly, 2008; Johanson & Brooks, 2010; Treece & Treece Jr, 1982;), represent slightly over 10% of the universe of students included in the main research study. This procedure allowed to ensure that the students' responses were not biased by a lack of comprehension of the questionnaire and also to prevent difficulties in deconstructing the answers to open-ended questions during the content analysis. Furthermore, it is important to highlight that in the final version of the measurement instrument, students were invited to rate the questionnaire regarding its objectivity and intelligibility, to guarantee that the questions were clear and well understood by all respondents.

Lastly, students from both the control and the experimental group rated the questionnaire as being objective and easy to understand, which further emphasizes the adequacy of the validated version of the questionnaire.

2.5. Data Analyses

Methods of descriptive and inferential statistics were used to analyze the pre-/post-test data. All statistical analyses were carried out using IBM's Statistical Package for the Social Sciences (SPSS) version 24.

Independent samples t-tests and paired samples t-tests for a 95% confidence interval were used for five-point Likert-type scale data and the effect size of mean differences registered with t-test was measured using Cohen's d (Cohen, 1988). Data gathered through open-ended questions and dichotomous variables were analyzed using Chi-square and the McNemar tests, respectively, and considering the phi coefficient as the effect size measure (Pallant, 2007). Furthermore, to obtain a broader, more inclusive depiction of the effectiveness of the activities, while strengthening the interpretation of the outcomes of the analyses performed (Punch, 2009), it was decided to combine quantitative and qualitative methods of analysis, as has been suggested in similar studies (Fonseca et al., 2012; Gelbart et al., 2009; Machluf & Yarden, 2013). This methodology would avoid missing detailed information that cannot be retrieved exclusively from quantitative data (Johnson & Christensen, 2012).

In what concerns the qualitative data, a thematic content analysis of the participants' responses to open-ended questions was performed with the purpose of producing a systematic description of the meaning of specific information gathered through the definition of coding categories (Schreier, 2012). This allowed to organize extensive answers to open-questions into fewer and more focused content categories (Hsieh & Shannon, 2005; Krippendorff, 2004; Weber, 1990). The analysis of the answers to the open-ended questions was performed according to the framework previously created by the authors in which specific categories of answers have been defined (Supplementary figure III - 1). Regarding the open-ended question Q9, aimed to assess students' interest, the subjective task value of Eccles *et al.* (Eccles, 2005; Eccles & Wigfield, 2002) that characterizes an expectancy-value model of achievement motivation was used as the theoretical framework underlying data analysis. Task value is related with the quality of the task, which influences the probability of it being select by an individual. In this study, the intrinsic/interest value (i.e. expected enjoyment of engaging in the task), the utility value (i.e. possible rewards from the task) and the cost of engaging in the activities were the dimensions considered when analyzing the students' answers.

3. Results and Discussion

3.1. Students' Literacy on Bioinformatics and its Applications

It is consensual that an updated and edifying high school level education requires an attentive revision of the curricula aligned with the challenges of NGSS and capable to meet Science, Technology, Engineering and Mathematics (STEM) education (Champagne Queloz et al., 2017; Kovarik et al., 2013; National Research Council, 2013; Wefer & Sheppard, 2008). In this regard, bioinformatics is in a privileged position, due to the transdisciplinary approach it entails, by seeking a level of integration of different disciplines, such as biology, computer science, and mathematics, beyond the mere interdisciplinary relationship between them. It is therefore reasonable to acknowledge the importance of integrating bioinformatics in high school, as emphasized in several studies (Dudley & Butte, 2009; Machluf et al., 2017; Machluf & Yarden, 2013; Magana et al., 2014; Marques et al., 2014), even though there is scarce research on how to do it (Campbell & Nehm, 2013; Machluf et al., 2017; Magana et al., 2014). To measure the impact of educational initiatives using bioinformatics resources on high school students, and to emend misconceptions and tailor adequate bioinformatics activities for successful learning, it is important to diagnose the knowledge students perceive to have about bioinformatics and genomics-related concepts (Champagne Queloz et al., 2017; Form & Lewitter, 2011; Gelbart et al., 2009; Gelbart & Yarden, 2006).

In the universe of 387 high school students enquired in the present study only a modest percentage (40.1% of experimental group; 24.2% of control group) revealed to have heard about bioinformatics in the pre-test (Q1), and most of the one who did so could not define bioinformatics, admitting that their answer reflected the etymological meaning of the word. Following an expositive teaching session on bioinformatics and associated resources, such as databases and applications, in the post-test, the percentage of the students who revealed to have heard about bioinformatics raised consistently for both the experimental group (99.0%), and the control group (99.0%) (Figure III - 4). Regardless the fact that in the post-test most of these students linked bioinformatics to the etymology of the word: bio + informatics (60.9% of experimental group; 73.6% of control group), which undermines a truly sensible diagnostic of their understanding of bioinformatics, some students did mention specific aspects, such as data analysis, storage and comparative genomics. The difference observed in this regard between the experimental and control groups (31.0% of experimental group; 22.0% of control group) may be explained by the fact that students in the experimental group carried out a set of bioinformatics exercises using the mentioned resources and used bioinformatics platform for comparative genomics, contrarily to their counterparts in the control group. This is particularly evident regarding comparative genomics, a completely new notion for the majority

of the students, which was mentioned by 6.6% of the experimental group students and only by 1.1% of the control group students. Furthermore, students from both groups recognized that gene banks are open-access resources (Q6.1) (81.2% of experimental group; 59.0% of control group), and generally accessible to all citizens (Q6.2) (78.8% of experimental group; 62.1% of control group), suggests an enhanced perception of what comprises a bioinformatics scientific toolbox and of their empowerment to access it (Figure III - 4). These findings, observed in other studies (Kovarik et al., 2013; Machluf et al., 2017), report for a motivational trigger of scientific literacy and STEM education. The higher percentage scores obtained with the experimental group indicate that complementing expositive teaching with hands-on *in silico* laboratories favors the acquisition of structural knowledge. This was a particularly relevant outcome that allows to dismiss the common misconception that bioinformatics analysis always requires programing skills. In fact, while initially, i.e. before the intervention, students from both groups (62.9% of experimental group; 69.5% of control group) agreed that programing skills would be needed to use bioinformatics tools (Q6.3) (Figure III - 4), after the intervention only 27.1% of the students from the experimental group and 32.6% of the control group agreed with this statement (Figure III - 4). These data indicate that while initially students associated bioinformatics analysis to a set of complex computer codes, after they were challenged with bioinformatics activities, they were able to acknowledge the panoply of bioinformatics applications with user-friendly interfaces tailored for web browsers that do not require programming competencies as has been highlighted by Martins, Lencastre, et al. (2018b). Students were shown to be aware that bioinformatics tools are essential to molecular biology studies (Q6.4), both in the pre- and post-test (Figure III - 4). Still, in the post-test there was a slight increment in the percentage of students who agree with this statement, suggesting that they confirmed their previous idea about the role of bioinformatics in molecular biology.

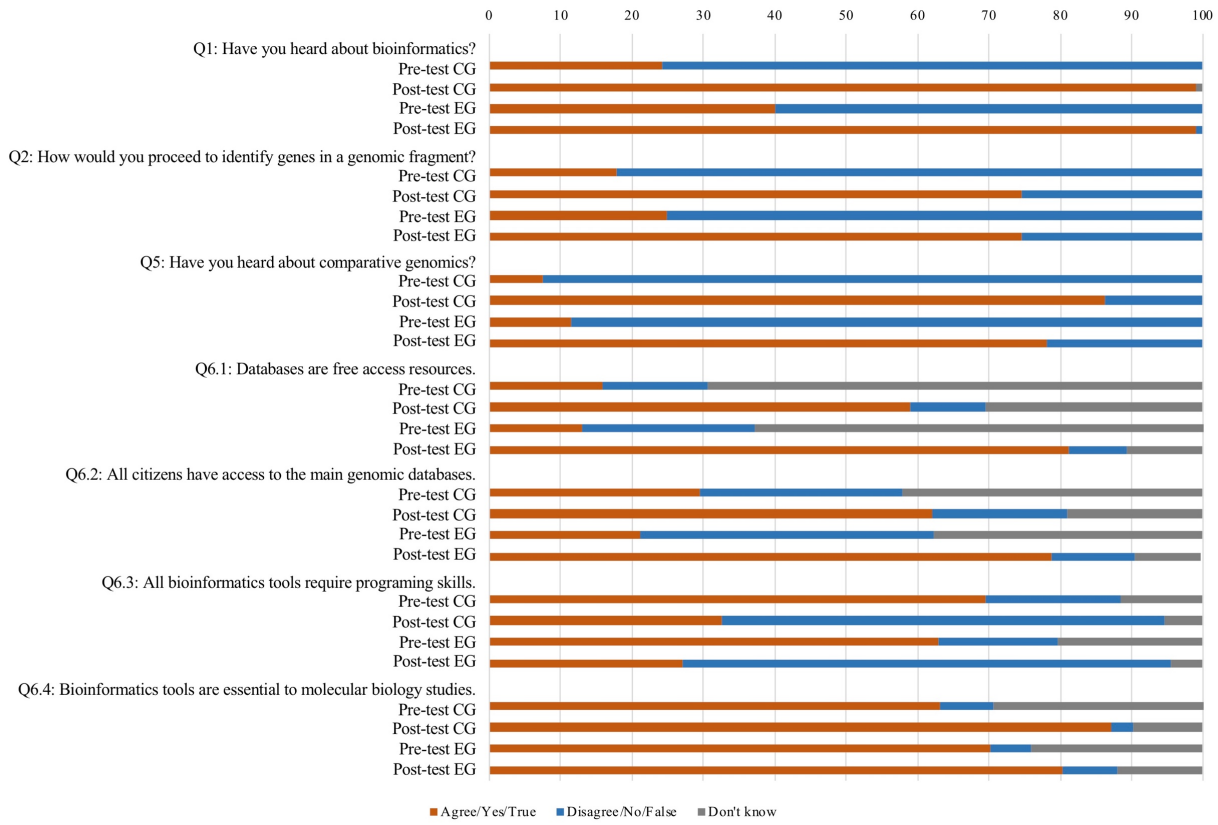


Figure III-4. Students' knowledge towards bioinformatics, gene regulation, and genomics.

Following the intervention (i.e. post-test), when the participants were asked to "*Indicate which bioinformatics platforms [they] used*" (Q8.5), 16.7% of students in the control group failed to mention any of the expected resources used during the intervention (namely NCBI; NCBI ORFfinder; NCBI BLAST; and MaGe). This percentage dropped to 1.7% in the experimental group (Figure III - 5), indicating the positive impact of bioinformatics laboratories on students' knowledge.

The bioinformatics exercises used in this study aimed to train the students on key procedures to identify genes from a genome sequence, as proposed by Martins, Fonseca, et al. (2018). Since the bioinformatics exercises were supported by a tutorial video comprising detailed guidelines and instructions (https://drive.google.com/file/u/1/d/1WrtjzLHzKI7nLALtVnmqy6WhPnkUsQSR/view), it was important to determine if the students' performance actually contributed to enhance their knowledge on basics genome mining, and did not resume to a mere mechanical procedure of following a recipe step by step. To address this question, the students were asked to describe the procedures that can be used to identify putative genes within a genomic DNA sequence (Q2; Q2.1). While in the pre-test only a minority of the students in both groups (24.9% of experimental group; 17.9% of control group) claimed to know the procedures to deconstruct a DNA sequence into putative coding sequences, in the post-test this percentage increased significantly (74.6% of experimental group; 74.7% of control group) (Figure III - 4). As

expected, the change between pre- and post-test is statistically significant for both groups (Supplementary Table III - 1). To fully elucidate if the students' perceptions were aligned with their knowledge, a content analysis was carried out.

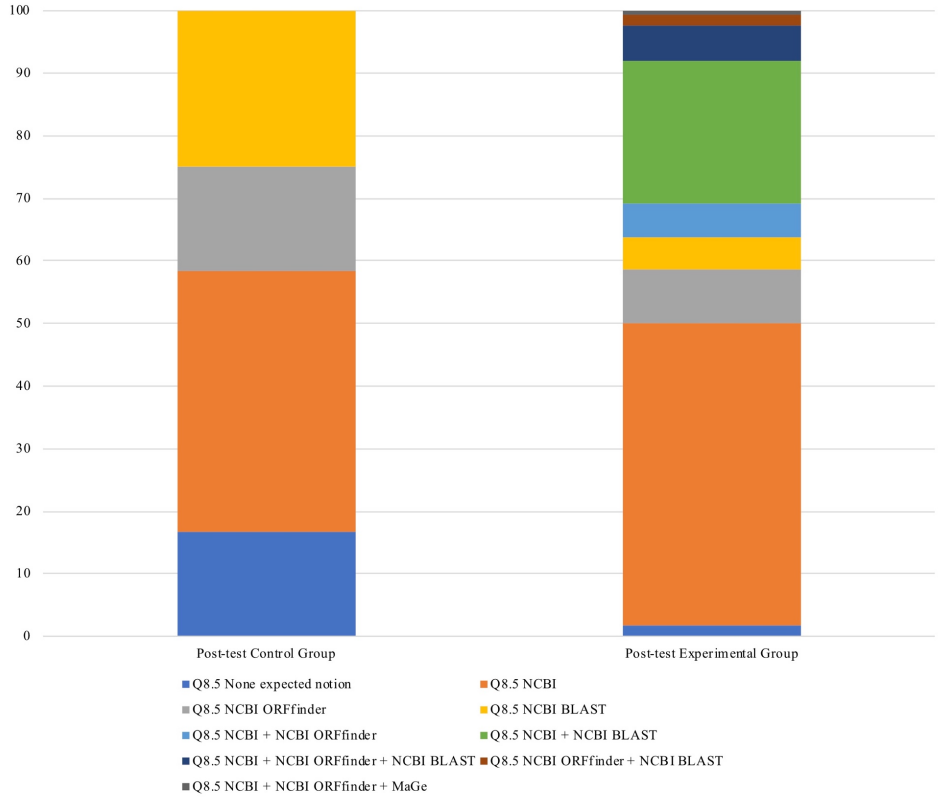


Figure III-5. Bioinformatics tools mentioned by students to unveil genes from bacterial genomics sequences.

In this regard, a framework with three expected bioinformatics-related notions was defined: 1) "*Getting the target DNA sequence in a database*"; 2) "*Looking for Open Reading Frames*"; 3) *"Deciding which of the retrieved ORFs are likely to be genes running a BLAST"*.

The pre-test content analysis regarding the answers to Q2.1 showed that students who admitted knowing how to identify putative genes from a genomic DNA sequence failed to mention any of the three notions. Instead, they mentioned, for instance, that "*To unveil a DNA sequence we can perform an electrophoresis to determine the genes, looking at the gel bands in comparison to a known gene. Restriction enzymes may be needed in this procedure*", which was the most frequently recorded notion in the experimental group, and that it is possible to "*Use the genetic code to identify the codons in a DNA sequence*", which was the most frequently recorded notion in the control group.

The post-test content analysis for the answers to Q2.1 revealed that 47.7% of students in the control group did not mention any of the expected answers, 9.0% mentioned one of the expected answers, 41.8% mentioned two expected notions and 1.5% mentioned all three

expected notions. This trend improved in the experimental group, for which the percentage of students who mentioned one expected notion (14.3%) and all the expected notions (11.1%) was higher. Furthermore, the percentage of students who did not mention expected notions was also lower in the experimental group than in control group (38.6%).

Contrary to what was observed in the pre-test, in the post-test students from both groups mentioned bioinformatics approaches, rather than wet lab techniques currently mentioned in their biology classes, such as electrophoresis and restrictions enzymes. This outcome highlights the notion that, following a bioinformatics laboratory, most of the experimental designs envisioned by students to address a research question are based on a bioinformatics approach, instead of involving wet lab techniques that were already known to them. More than suggesting an enrichment of students' scientific toolbox and the development of thinking skills, the intervention seems to narrow the gap between students' school reality and what are common research practices nowadays, which is consistent with the educational benefits of bioinformatics reported in the literature (Flanagan, 2013; Gelbart & Yarden, 2006; Wood & Gebhardt, 2013). The data further suggests that when students are guided in the use of a wide variety of resources they show to be capable to explore ideas and to interpret results in order to answer questions raised by the teacher (Kuhlthau et al., 2007).

3.2. Students' Knowledge on Gene Regulation and Genomics

According to the educational theories proposed by Ausubel (1968) and Vygotskiĭ & Cole (1978), students' prior knowledge, and in particular students' misconceptions, is of crucial importance when learning a new issue. Several diagnostic instruments are available, in published research studies, that can be used to obtain guidelines for specific interventions to address these misconceptions (Gurel et al., 2015; Klymkowsky et al., 2010; Tsui & Treagust, 2010). Examples of students' key misconceptions regarding basic genetic and genomics notions are already described in the literature, and include the use of gene and genome as synonyms; the misunderstanding of the relationship between a gene and DNA; a misinterpretation of the association between a gene and gene regulation; and the idea that some organisms as bacteria and fungi, often do not have DNA (Lewis & Kattmann, 2004; Shaw et al., 2008). Adding the relevance of addressing these misconceptions, the Portuguese biology curriculum for the 11th grade (Mendes et al., 2003) recommends the discussion the concept of "*codogene*" - part of a gene, i.e., a triplet of DNA -, which is contributing to mislead students on the definition of gene. Having in mind the reported misconceptions, the activities implemented in this study aimed to tackle notions related with genes, genomes, alternative start codons and the genetic code. Participants of both groups agreed that different bacteria groups have genes in common (Q7.1) and were shown to be aware that not all bacteria genes are identified and characterized, and that genomics information is still missing for many

species (Q7.4) (Figure III - 6). Conversely, misconceptions related with gene structure and the features of the genetic code did not seem to be overcome following the activities. In fact, students of both groups tended to disagree with the existence of different start codons (Q7.2) and were shown to be unaware of the existence of a bacterial genetic code (Q7.3), both in pre- and post-test (Figure III - 6).

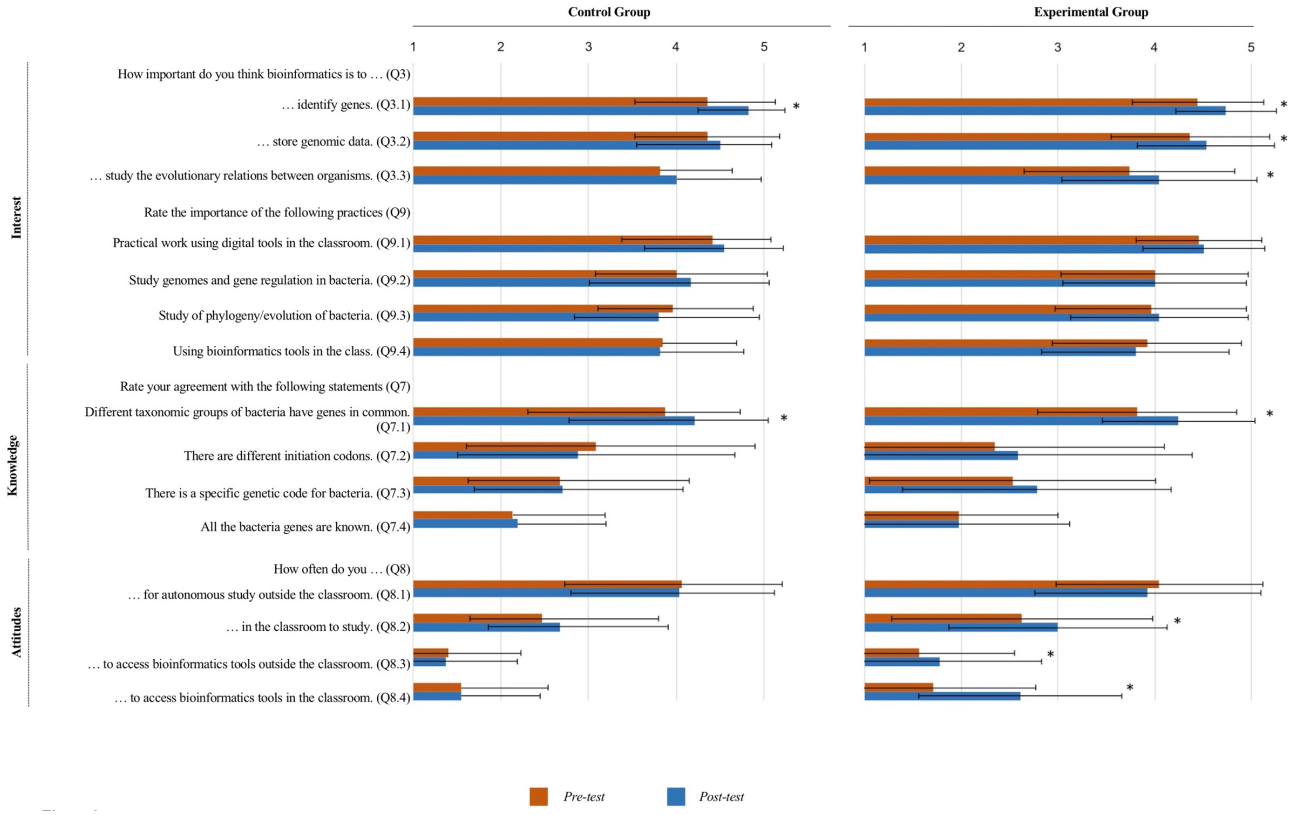


Figure III-6. Students' knowledge, interest, and attitudes towards the integration of bioinformatics in science curricula.

These two questions were conceived having in mind that in high school it is commonly taught that there is a unique start codon, a misconception that is reinforced in most textbooks. During the practical activities, students from the experimental group explored different start codons and worked with a specific bacteria-dedicated genetic code when using the tool NCBI ORFfinder, which was expected to make them aware of the specifications of the genetic code. However, surprisingly, the acquisition of this knowledge was not confirmed which can be explained by reported evidence that even after being taught and accurately updated on a given scientific content for which misconceptions are observed, many students do not reconstruct their thinking (Shaw et al., 2008). In this study, the practical component designed to address this particular misconception was also not effective. In fact, the use of misleading terms, simplified explanations that induce erroneous interpretations, adapted language and everyday examples to explain biological phenomena is often the origin of students' misconceptions, which can be tenacious and quite difficult to be overcome, ending up being perpetuated all through their high school education (Cho et al., 1985; Shaw et al., 2008; Soyibo, 1995;

Tekkaya, 2003). These data call for further attention and suggest that exercises specifically dedicated to exploring different start codons and distinct genetic codes according to the taxa of interest, are needed to successfully overcome these deep-rooted misconceptions.

Other knowledge dimensions analyzed in this study include the concepts of genomics (Q4) and comparative genomics (Q5) aimed to acknowledge the importance of genomics in nowadays science and how it is impacting common societal sectors such as human health and biotechnology. The results recorded for these two questions (Q4 and Q5) revealed a noticeable lack of knowledge about these concepts as previously described (Baumler et al., 2012; Chen & Kim, 2014; Kirkpatrick et al., 2002; Shaw et al., 2008), which bares implications when trying to use bioinformatics tools.

In the post-test, 54.7% of the students in the experimental group provided a correct definition of genomics, i.e. "*The field of science that studies genomes"*, trend that was not registered in the control group in which only 29.1% of the students were able to define this concept correctly. Zooming in the answers to identify the reported misuse of gene and genomics in an interchangeable way, evidences a significant difference between the control and the experimental groups. In the pre-test, 1.5% of control group students mentioned that genomics is a field of science that studies genes and/or genomes, a frequency which increased in the post-test (5.1%). In turn, in the experimental group, the trend was opposite, with the frequency of these notions decreasing from the pre- to the post-test (8.2% *vs.* 0.5%). These differences suggest an improvement of the quality of the answers of the students who carried out the bioinformatics exercises, i.e. the experimental group, apparently denoting that the expository teaching failed to clearly teach the difference between genomics and genetics. This may have resulted in the lack of accuracy witnessed in students' replies to question Q4, in what relates to the reference genome instead of gene. It is important to mention that in the particular case of the Portuguese science curriculum and in the NGSS, genomics is not at all mentioned; the topic addressed when referring to gene and genome-related issues is genetics. In this regard, before the intervention, only a few students mentioned that they had heard about comparative genomics (Q5) (Figure III - 4), an important concept that currently is not addressed in science classes (Martins & Tavares, 2018).

When students were asked to define comparative genomics in the post-test (Q5.1), the majority was able to do so correctly (79.5% in the experimental group; 75.3% in the control group). They associated the field with "*genomic characteristics/genomes/genes/DNA sequences/homologous between different organisms*", which suggests that the expository teaching on comparative genomics was efficient in fostering an accurate understanding about comparative genomics in students in both groups. As comparative genomics was a notion new to students, it was not conditioned by their previous perceptions, contrary to what happened

with the concepts of genetics and genomics. Despite this general trend, question 5.1 was also aimed to depict more misconceptions that could be associated with the definition of comparative genomics. In the pre-test, 3.6% of the students in the experimental group mentioned that comparative genomics could be defined as comparisons between genes and the phenotype, claimed that comparative genomics is the comparison between genetic sequences. The percentages of students with these misconceptions in the experimental and control groups lowered significantly in the post-test (2.4% and 1.4%, respectively). At this stage, i.e. in the post-test, a new notion was identified, with experimental group students associating comparative genomics with the "*Comparison of genomes of two or more species aiming to investigate phylogenetic relations*" (6.7%). Having these outcomes in mind, it can be noted the quality of answers of students in the experimental group improved after the intervention. It is worth mentioning that in the post-test 5.2% of control group students also recognized that comparative genomics can be associated with phylogenetic studies, which can be justified by the expositive teaching.

3.3. Attitudes and Interest

Together with the characterization of the students' knowledge regarding bioinformatics, gene regulation and genomics, as described in the previous section, a depiction of their attitudes and interest towards bioinformatics was also carried out. As previously mentioned, in the context of this study, interest was interpreted according to Eccles expectancy-value model (Eccles, 2005), which foresees motivation as a result of the combination of expectancy and value. The value given by students to a specific task is extremely important because they are more likely to pursue an activity if they acknowledge it worth. The model further differentiates task value into four components: attainment value (importance of doing it correctly), intrinsic value (personal enjoyment), utility value (perceived usefulness for future goals), and cost (competition with other goals) (Eccles, 2005; Eccles & Wigfield, 2002; Leaper, 2011).

From the start, students were shown to be aware about the importance of bioinformatics to identify genes (Q3.1). Nevertheless, the classroom discussion that followed the expository teaching session about the need of bioinformatics tools to efficiently mine the huge genomics datasets, contributed to reinforce this belief as demonstrated by the statistically significant difference between pre- and post-test results (Figure III - 6, Table III - 1).

Regarding the role of bioinformatics to store genomic data (Q3.2) and to study evolution (Q3.3), a statistically significant difference was observed from pre- to post-test in the experimental group (Figure III - 6, Table III - 1), but not in the control group. As the bioinformatics laboratories entailed the recruitment of bioinformatics resources particularly suited to access large data sets and address evolutionary inferences through synteny maps,

these results highlight the direct impact of the intervention, which sustains identical results detailed in other studies (Kremer et al., 2005; Luscombe et al., 2001).

When asked to rate the importance of studying gene regulation (Q9.2) and evolution in bacteria (Q9.3), students in both groups agreed on its importance in both assessment moments (Figure III - 6). In what concerns the study of gene regulation (Q9.2.1), in the control group, its perceived importance was mainly connected with its usefulness from an instrumental point of view (60.3%) (utility value), as suggested by expressions that linked its importance with the goals such as: *"To get in touch with the world around us"* or *"To improve human life quality"*. Interestingly, in the experimental group, adding to the utilitarian value (42.7%), a more knowledge-related intrinsic worth (intrinsic value) was also well represented (42.7%), as shown by statements such as "*When we study bacteria it is interesting to have the chance to better understand this group and to get information about their metabolism in different environments*". These results indicate that the scientific topic chosen for these activities is of interest to the students, and that the bioinformatics exercises carried out by the experimental group contributed to a more focused appraisal of the relevance of genomics and gene regulation. An identical trend was observed concerning the interest of evolutionary studies in bacteria (Q9.3.1), with 59.0% of students in the experimental group and 50.9% of students in the control group mentioning notions that reflect their motivation to explore the scientific topic, which emphasizes the importance of adding comparative genomic tools to the activities proposed.

Table III-1. Pre- and post-test comparison of students' knowledge, interest, and attitudes towards bioinformatics.

		Control Group				Experimental Group			
		t	*df*	*p*	|d|	*t*	*df*	*p*	|d|
Interest	*How important do you think bioinformatics is to ... (Q3)*								
	... identify genes. (Q3.1)	-6.27	89	<0.01*	0.77	-5.94	274	<0.01*	0.48
	... store genomic data. (Q3.2)	-1.59	90	0.12	0.21	-2.66	271	0.01*	0.21
	... study the evolutionary relations between organisms. (Q3.3)	-1.87	78	0.07	0.22	-3.62	230	<0.01*	0.30
	Rate the importance of the following practices (Q9)								
	Practical work using digital tools in the classroom. (Q9.1)	-1.78	94	0.08	0.18	-1.32	281	0.19	0.08
	Study of genomes and gene regulation in bacteria. (Q9.2)	-1.65	72	0.10	0.17	0.07	209	0.94	0.00
	Study of phylogeny / evolution of bacteria. (Q9.3)	0.48	62	0.63	0.15	-1.64	221	0.10	0.09
	Using bioinformatics tools in the class. (Q9.4)	0.00	41	1.00	0.03	1.33	181	0.18	0.12
Knowledge	*Rate your agreement with the following statements (Q7)*								
	Different taxonomic groups of bacteria have genes in common. (Q7.1)	-4.12	66	<0.01*	0.40	-6.08	182	<0.01*	0.47
	There are different initiation codons. (Q7.2)	0.99	92	0.33	0.11	-1.78	279	0.08	0.13
	There is a specific genetic code for bacteria. (Q7.3)	0.32	65	0.75	0.02	-1.61	191	0.11	0.17
	All the bacterial genes are known. (Q7.4)	-0.22	66	0.83	0.05	0.00	207	1.00	0.00
Attitudes	*How often do you use the computer/technological devices... (Q8)*								
	... for autonomous study outside the classroom. (Q8.1)	0.20	94	0.84	0.02	1.72	290	0.09	0.11
	... in the classroom to study. (Q8.2)	-1.86	94	0.07	0.16	-5.03	290	<0.01*	0.30
	... to access bioinformatics tools outside the classroom. (Q8.3)	0.28	94	0.78	0.04	-2.82	285	0.01*	0.20
	... to access bioinformatics tools in the classroom. (Q8.4)	0.00	94	1.00	0.00	-11.89	286	<0.01*	0.85

t – paired samples t-test for a 95% confidence interval (*p*). *df* – degrees of freedom. *d* – Cohen's *d* measure of effect size. (*) indicates significant differences between pre- and post-test to each group.

As expected, students considered the practical work using digital tools important, engaging and motivating, raising their intrinsic interest (Q9.1; Q9.1.1) (Figure III - 6). Concerning students' interest on the use of bioinformatics tools in the classroom, even before the *in silico* laboratories, they had already shown to be motivated in this regard (Q9.4) (Figure III - 6). Despite the lack of statistically significant differences (Table III - 1), in the post-test the students from both groups agreed that the integration of bioinformatics laboratories in the classroom (Q9.4) can have a beneficial impact to increase their intrinsic interest. This suggests their curiosity and awareness about the potential of using these tools in the classroom, regardless of whether they carried out (experimental group) or not (control group) the bioinformatics exercises.

Interesting remarks on the participants' engagement and interaction can be made based on the observations carried out during the implementation of the activities. For instance, the students were very surprised when they realized the incredible amount of open-access biological data, as translated by questions of amazement like: "*Can I access these bioinformatics resources for free at home?"; and "Nice! Everyone can do it?*". Having in mind we are now living in the post-genomic era, these reflections are crucial for students to get acquainted with genomics data sharing and to become aware of the social benefits and ethical implications of open access data (Foster & Sharp, 2007; Oliver et al., 2012).

Another aspect that students stated as being truly interesting pertained to the fact that they were sharing the exact same platforms used by professional researchers. These findings meet the reported importance of exposing science students to real-world phenomena and data, since this kind of activities can increase their interest and better prepare them for engaging in careers in science (Flanagan, 2013; Gelbart & Yarden, 2006). Furthermore, the observations showed that after completing the activities, students looked forward to exploring other tools in the platforms suggested, making comments such as: "*What is the size of the genome of a spider*?"; "*Are virus - such as HIV-, genomes also available at this database*?"; or "*Let us search for the gene coding for insulin*." While this enhanced enthusiasm and curiosity have been reported for university science students (Chapman et al., 2006; Madlung, 2018), it has been poorly described in pre-university levels of education, which makes this finding even more interesting.

Confirming the participants' interest in learning science with bioinformatics tools is the fact that only a low percentage of students (13.5% in the experimental group; 9.3% in the control group) associated the integration of bioinformatics in the class (Q9.4.1) with a cost, according to Eccles Framework (Eccles, 2005). These students mentioned that incorporating bioinformatics in the classroom "*is not that important once there are similar ways of obtaining the same results*"; or that "*According to the Portuguese curricula for science in high school*

there is no need of using such complex tools"; and also "*This kind of activities can make classes more confusing since students are not used to working with these applications*". These comments seem to reveal a lack of sympathy for innovative learning challenges.

As it is well-known, nowadays youths are particularly at ease with digital resources (Quinn & Oldmeadow, 2013; Rosen et al., 2010) and, indeed, students from both the experimental and the control groups admitted that they often take advantage of the technologies at their disposal in their autonomous study outside the classroom (Q8.1) (Figure III - 6). Despite this reality, students from both groups stated that they do not use computers or other technological devices in the classroom (Q8.2, Figure III - 6). The statistically significant pre- to post-test increase observed in the answers to this question among the experimental group students is likely due to the unique opportunity created by this study for them to join bioinformatics laboratories (Table III - 1). Recent studies reported that although schools apparently have the necessary conditions to successfully integrate Information and Communications Technology (ICT) in the classroom, there are still barriers, such as teachers' pedagogical beliefs, which prevent the use of computers in classroom settings (Ertmer, 2005; Marcinkiewicz, 1993; Sang et al., 2010). Interestingly some informal comments made by the students revealed that their teachers often feel discouraged to use technology in the classroom because they do not feel comfortable with it, which meets the constraints mentioned by the teacher, the majority of whom acknowledged their anxiety regarding the use of technology in this setting (Machluf & Yarden, 2013; Martins, Lencastre, et al., 2018a, 2020).

Even though students of both groups also revealed (Q8.3) that they usually do not access bioinformatics tools outside the classroom, there is a significant pre- to post-test difference for the experimental group, which may suggest that these students decided to take advantage of the bioinformatics resources explored after the activities (Figure III - 6; Table III - 1). Regarding the specific use of bioinformatics tools in the classroom (Q8.4), while in the pre-test students from both groups answered negatively to this question, as expected, in the post-test the students from the experimental group reported that they used bioinformatics in their classes (Figure III - 6; Table III - 1).

Having in mind that the students who took part in this study belong to a highly technological society, one can anticipate that their performance in manipulating computer-based tools was efficient (Quinn & Oldmeadow, 2013; Rosen et al., 2010). Indeed, and regardless that most of the students had never experienced working with bioinformatics tools before, during the implementation of the bioinformatics laboratories no major difficulties to follow the guidelines and discussing the issues raised were reported to the teacher. The observations showed that students were completely able to manage the platforms and did not feel the need to use printed out guidelines. Instead, they looked for solutions and alternatives together with their

classmates and took advantage of the technological resources available, namely smartphones. In spite of the expectable side talk, the participants' behavior and their questions and comments suggest their engagement in every task that they were asked to perform.

4. Conclusion

The findings obtained in this study demonstrate an improvement in students' knowledge of concepts such as gene, protein synthesis, nucleic acid (DNA, RNA), start and stop codons, genome, evolutionary relations, and genomic or comparative genomics, following their participation in bioinformatics-based activities "*Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Students Discovery of Genes*" (Martins, Fonseca, et al., 2018). By the end of the activities, students were also shown to be more aware of the applications and potential of bioinformatics.

This study also raises several questions that are worth pursuing in future research namely related with misconceptions that were addressed in this intervention. In addition, future focus on other school levels (namely middle school) and other curricular topics might be relevant to cross-examine and more widely and consistently depict the impact of bioinformatics-based activities in the classroom. Likely pertinent will be to assess the influence of the "*teacher*" in students' performance through a nested effect analysis.

Beyond the evidence of the educational benefits of incorporating practical activities in science education programs, overall this study represents a contribution to introduce a top-notch research area – bioinformatics – in school and to inform stakeholders about its potential from not only educational but also scientific and other social points of view.

Data Availability Statement

The raw data supporting the conclusions of this article will be made available by the authors, without under reservation.

Ethics Statement

The project was institutionally approved by each school's Directive Board. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author Contributions

AM, MJF and FT designed the research plan. AM followed the implementation of the instruments in the classroom and collected, organized and analyzed the data. MJF, ML, LL and FT participated in data analysis and interpretation. All authors contributed to the writing of this article and approved the submitted version.

Funding

AM is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT [SFRH/BD/112038/2015]. This work is supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia in the scope of the project UIDB/50027/2020.

Acknowledgments

The authors are grateful to all participants (teachers, students, and schools) for creating the opportunity to implement these bioinformatics laboratories and carry out this study.

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.578099/full#supplementary-material and at the end of this manuscript in this thesis.

The files sent for reviewers' appraisal only, are made available in this thesis at the end of this publication.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Amenkhienan, E., & Smith, E. J. (2006). A web-based genetic polymorphism learning approach for high school students and science teachers. *Biochemistry and Molecular Biology Education*, *34*(1), 30–33. https://doi.org/10.1002/bmb.2006.49403401030

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., & Schneider, M. V. (2017). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*. https://doi.org/10.1093/bib/bbx100

Ausubel, D. P. (1968). *Educational psychology: a cognitive view*. Holt, Rinehart and Winston.

Baumler, D. J., Banta, L. M., Hung, K. F., Schwarz, J. A., Cabot, E. L., Glasner, J. D., & Perna, N. T. (2012). Using Comparative Genomics for Inquiry-Based Learning to Dissect Virulence of *Escherichia coli* O157:H7 and *Yersinia pestis*. *CBE—Life Sciences Education*, *11*(1), 81–93. https://doi.org/10.1187/cbe.10-04-0057

Campbell, C. E., & Nehm, R. H. (2013). A critical analysis of assessment quality in genomics and bioinformatics education research. *CBE Life Sciences Education*, *12*(3), 530–541. https://doi.org/10.1187/cbe.12-06-0073

Champagne Queloz, A., Klymkowsky, M. W., Stern, E., Hafen, E., & Köhler, K. (2017). Diagnostic of students' misconceptions using the Biological Concepts Instrument (BCI): A method for conducting an educational needs assessment. *PLoS ONE*, *12*(5), e0176906. https://doi.org/10.1371/journal.pone.0176906

Chapman, B. S., Christmann, J. L., & Thatcher, E. F. (2006). Bioinformatics for undergraduates: Steps toward a quantitative bioscience curriculum. *Biochemistry and Molecular Biology Education*, *34*(3), 180–186. https://doi.org/10.1002/bmb.2006.49403403180

Chen, L.-S., & Kim, M. (2014). Needs Assessment in Genomic Education. *Health Promotion Practice*, *15*(4), 592–598. https://doi.org/10.1177/1524839913483470

Cho, H. -H, Kahle, J. B., & Nordland, F. H. (1985). An investigation of high school biology textbooks as sources of misconceptions and difficulties in genetics and some suggestions for teaching genetics. *Science Education*, *69*(5), 707–719. https://doi.org/10.1002/sce.3730690512

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates.

Connelly, L. M. (2008). Pilot studies. *MedSurg Nursing*, *17*(6), 411–413.

Dudley, J. T., & Butte, A. J. (2009). A Quick Guide for Developing Effective Bioinformatics Programming Skills. *PLoS Computational Biology*, *5*(12), e1000589. https://doi.org/10.1371/journal.pcbi.1000589

Eccles, Jacqueline S. (2005). Subjective task value and the Eccles et al. model of achievement-related choices. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of Competence and Motivation* (pp. 105–121). The Guilford Press.

Eccles, Jacquelynne S, & Wigfield, A. (2002). Motivational beliefs, values and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development*, *53*(4), 25–39. https://doi.org/10.1007/BF02504683

Flanagan, J. (2013, May). *Open data for science education*. PLoS Blogs. https://doi.org/10.1525/bio.2010.60.5.2

Fonseca, M. J., Costa, P., Lencastre, L., & Tavares, F. (2012). Multidimensional analysis of high-school students' perceptions about biotechnology. *Journal of Biological Education*, *46*(3), 129–139. https://doi.org/10.1080/00219266.2011.634019

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. In *PLoS Computational Biology* (Vol. 7, Issue 10). https://doi.org/10.1371/journal.pcbi.1002243

Foster, M., & Sharp, R. (2007). Share and share alike: deciding how to distribute the scientific and social benefits of genomic data. *Nature Reviews Genetics*, *8*(8), 633–639. https://doi.org/10.1038/nrg2124

Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research simulation in bioinformatics. *Journal of Biological Education*, *40*(3), 107–112. https://doi.org/10.1080/00219266.2006.9656026

Gelbart, Hadas, Brill, G., & Yarden, A. (2009). The Impact of a Web-Based Research Simulation in Bioinformatics on Students' Understanding of Genetics. *Research in Science Education*, *39*(5), 725–751. https://doi.org/10.1007/s11165-008-9101-1

Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, *11*(5), 989–1008. https://doi.org/10.12973/eurasia.2015.1369a

Hsieh, H.-F., & Shannon, S. E. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research*, *15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Johanson, G. A., & Brooks, G. P. (2010). Initial Scale Development: Sample Size for Pilot Studies. *Educational and Psychological Measurement*, *70*(3), 394–400. https://doi.org/10.1177/0013164409355692

Johnson, B., & Christensen, L. B. (2012). *Educational research : quantitative, qualitative, and mixed approaches*. SAGE Publications.

Kirkpatrick, G., Orvis, K., & Pittendrigh, B. (2002). A teaching model for biotechnology

and genomics education. *Journal of Biological Education*, *37*(1), 31–35. https://doi.org/10.1080/00219266.2002.9655843

Klymkowsky, M. W., Underwood, S. M., & Garvin-Doxas, R. K. (2010). *Biological Concepts Instrument (BCI): A diagnostic tool for revealing student thinking*. http://arxiv.org/abs/1012.4501

Koch, I., & Fuellen, G. (2008). A review of bioinformatics education in Germany. *Briefings in Bioinformatics*, *9*(3), 232–242. https://doi.org/10.1093/bib/bbn006

Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Kremer, A., Schneider, R., & Terstappen, G. C. (2005). A Bioinformatics Perspective on Proteomics: Data Storage, Analysis, and Integration. *Bioscience Reports*, *25*(1–2), 95–106. https://doi.org/10.1007/s10540-005-2850-4

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology*. SAGE Publications.

Kuhlthau, C. C., Caspari, A. K., & Maniotes, L. K. (2007). *Guided inquiry : learning in the 21st century*. Libraries Unlimited.

Leaper, C. (2011). More Similarities than Differences in contemporary Theories of social development?. A plea for theory bridging. *Advances in Child Development and Behavior*, *40*, 337–378. https://doi.org/10.1016/B978-0-12-386491-8.00009-8

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education*, *26*(2), 195–206. https://doi.org/10.1080/0950069032000072782

Luscombe, N., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*. https://doi.org/10.1053/j.ro.2009.03.010

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible—the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660.

https://doi.org/10.1093/bib/bbt030

Madlung, A. (2018). Assessing an effective undergraduate module teaching applied bioinformatics to biology students. *PLOS Computational Biology*, *14*(1), e1005872. https://doi.org/10.1371/journal.pcbi.1005872

Magana, A. J., Taleyarkhan, M., Alvarado, D. R., Kane, M., Springer, J., & Clase, K. (2014). A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE—Life Sciences Education*, *13*(4), 607–623. https://doi.org/10.1187/cbe.13-10-0193

Marcinkiewicz, H. R. (1993). Computers and Teachers. *Journal of Research on Computing in Education*, *26*(2), 220–237. https://doi.org/10.1080/08886504.1993.10782088

Marques, I., Almeida, P., Alves, R., Dias, M. J., Godinho, A., & Pereira-Leal, J. B. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2020). Bioinformatics, a Befitting Tool for e-learning: Potential and Constraints according Teachers' Perceptions. In M. F. Costa & J. B. Dorrío (Eds.), *Hands-on Science. Science Education. Discovering and understanding the wonders of Nature* (pp. 97–105). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2018a). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. *3rd International Conference on Teacher Education (INCTE)*, 203–214. http://hdl.handle.net/10198/17381

Martins, A., Lencastre, L., & Tavares, F. (2018b). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 309–317). Hands-on Science Network.

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In Manuel Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

Mendes, A., Rebelo, D., & Pinheiro, E. (2003). *Programa de Biologia e Geologia 11º ou 12ºano(s)*. Ministério da Educação: Departamento do Ensino Secundário.

Mendes, A., Rebelo, D., & Pinheiro, E. (2004). *Biologia 12ºano - Curso Científico Humanístico de Ciências e Tecnologias*. Ministério da Educação.

National Research Council. (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

NCBI Resource Coordinators (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Newman, L., Duffus, A. L. J., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

Oliver, J., Slashinski, M., Wang, T., Kelly, P., Hilsenbeck, S., & McGuire, A. (2012). Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives. *Public Health Genomics*, *15*(2), 106–114. https://doi.org/10.1159/000334718

Pallant, J. (2007). *SPSS - Survival Guide to Data Analysis using SPSS for Windows*. Open University Press/McGraw-Hill.

Punch, K. (2009). Introduction to research methods in education. In *Introduction to research methods in education*.

Quinn, S., & Oldmeadow, J. A. (2013). Is the iGeneration a "we" generation? Social networking use among 9- to 13-year-olds and belonging. *The British Journal of Developmental Psychology*, *31*(Pt 1), 136–142. https://doi.org/10.1111/bjdp.12007

Rosen, L. D., Carrier, M. L., & Cheever, N. A. (2010). *Rewired : understanding the iGeneration and the way they learn*. Palgrave Macmillan.

Sadek, H. A. (2004). *Bioinformatics : principles, basic internet applications*. Trafford Publishing.

Sang, G., Valcke, M., Braak, J. van, & Tondeur, J. (2010). Student teachers' thinking processes and ICT integration: Predictors of prospective teaching behaviors with educational technology. *Computers & Education*, *54*(1), 103–112. https://doi.org/10.1016/J.COMPEDU.2009.07.010

Schreier, M. (2012). *Qualitative Content Analysis in Practice.* SAGE Publications.

Shaw, K., Horne, K., Zhang, H., & Boughman, J. (2008). Essay contest reveals misconceptions of high school students in genetics content. In *Genetics* (Vol. 178, Issue 3, pp. 1157–1168). Genetics Society of America.

https://doi.org/10.1534/genetics.107.084194

Soyibo, K. (1995). A Review of Some Sources of Students' Misconceptions in Biology. *Singapore Journal of Education*, *15*(2), 1–11. https://doi.org/10.1080/02188799508548576

Taylor, J. M., Davidson, R. M., & Strong, M. (2014). Drug-resistant tuberculosis: A genetic analysis using online bioinformatics tools. *American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Tekkaya, C. (2003). Remediating high school students' misconceptions concerning diffusion and osmosis through concept mapping and conceptual change text. *Research in Science and Technological Education*, *21*(1), 5–16. https://doi.org/10.1080/02635140308340

Treece, E. W., & Treece Jr, J. W. (1982). *Elements of research in nursing*. St. Louis : Mosby.

Tsui, C., & Treagust, D. (2010). Evaluating Secondary Students' Scientific Reasoning in Genetics Using a Two-Tier Diagnostic Instrument. *International Journal of Science Education*, *32*(8), 1073–1098. https://doi.org/10.1080/09500690902951429

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fèvre, F., Longin, C., Mornico, D., Roche, D., Rouy, Z., Salvignol, G., Scarpelli, C., Thil Smith, A. A., Weiman, M., & Médigue, C. (2013). MicroScope--an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research*, *41*(Database issue), D636-47. https://doi.org/10.1093/nar/gks1194

Vygotskiĭ, L. S. (Lev S., & Cole, M. (1978). *Mind in society : the development of higher psychological processes*. Harvard University Press.

Weber, R. (1990). *Basic Content Analysis*. SAGE Publications. https://doi.org/10.4135/9781412983488

Wefer, S. H., & Sheppard, K. (2008). Bioinformatics in High School Biology Curricula: A Study of State Science Standards. *CBE—Life Sciences Education*, *7*(1), 155–162. https://doi.org/10.1187/cbe.07-05-0026

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School—New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Supplementary figure III-1. Open-ended questions and answers' categorization system used to perform content analysis regarding knowledge dimension.

Question	**Q1.1:** Describe what is bioinformatics for you.	**Q2.1:** Indicate the main procedures that would follow to identify the gene(s) present in that sequence.	**Q4:** What is genomics for you?	**Q5.1:** Define comparative genomics.
Intervention Aim	Keep students in touch with the emerging field of bioinformatics.	Highlight the procedures needed to identify a gene in a genomic sequence using bioinformatics tools.	Teach the definition of genomics.	Teach the definition of comparative genomics.
Question Aim	Diagnose the students' conceptions about bioinformatics.	Appraise the procedures students describe to identify a gene in a genomic sequence.	Diagnose the students' conceptions about genomics.	Diagnose the students' conceptions about comparative genomics.
Answers' categorization system	⇒ Etymology ⇒ Applications: → Data Analysis → Data Storage → Comparative Genomics ⇒ Other conceptions: → Technology → Learning tool → Interdisciplinary	⇒ Expected Answer *(bioinformatics tools)* → Get a DNA sequence from a database → Looking for Open Reading Frames using ORFfinder tool → Running a BLAST of different ORF's. ⇒ Other procedures (examples) → Performing an electrophoreses to determine the genes → Looking at the gel bands and comparing with a reference gene → Restriction enzymes can be needed	⇒ Expected Answer → Science that studies the genomes. ⇒ Misconceptions → Bioinformatics is synonym of genomics. → Gene and genome are the same/students use the terms indifferently. → Genetics (study of genes) is the same that genomics (study of genes). ⇒ Other conceptions	⇒ Expected Answer → Genomic characteristics/genomes/ genes/DNA sequences/comparison of homologous. ⇒ Misconceptions → Comparison of genes and their phenotypic product. → Comparison between specific genetic sequences. ⇒ Other conceptions

Supplementary table III-1. Comparison between pre- and post-test students' knowledge towards bioinformatics, gene regulation and genomics.

		Control Group (Pre-test vs. Post-test)			**Experimental Group** (Pre-test vs. Post-test)		
		n	*p*	*Phi*	n	*p*	*Phi*
Knowledge	Q1: Have you heard about bioinformatics?	94	-	-	292	<0.01*	0.08
	Q2: Imagine the following situation: "As a researcher, you sequence a genomic fragment. Do you have any idea how you would proceed to identify the gene (s) present?"	95	<0.01*	0.21	281	<0.01*	0.20
	Q5: Have you heard about comparative genomics?	94	<0.01*	0.11	289	<0.01*	0.08
	Q6.1: Databases are free access resources.	25	0.04*	0.32	105	<0.01*	0.19
	Q6.2: All citizens have access to the main genomic databases.	52	0.02*	0.28	165	<0.01*	0.14
	Q6.3: All bioinformatics tools require programming skills.	80	<0.01*	0.21	223	<0.01*	0.08
	Q6.4: Bioinformatics tools are essential to molecular biology studies.	62	0.22	0.25	199	0.52	0.21

n – number of participants; McNemar test for a 95% confidence interval; Phi – Phi coefficient measure of effect size. (*) indicates significant differences between pre- and post-test to each group.

Supporting information:

These supplementary files were sent for reviewers' appraisal only.

Supplementary table III-2. Comparison between control and experimental groups' knowledge towards bioinformatics, gene regulation and genomics.

		Pre-test (Control *vs* Experimental Group)			**Post-test** (Control *vs* Experimental Group)		
		n	*df*	*p*	n	*df*	*p*
Knowledge	Q1: Have you heard about bioinformatics?	387	1	<0.01*	386	1	0.32
	Q2: Imagine the following situation: "As a researcher, you sequence a genomic fragment. Do you have any idea how you would proceed to identify the gene (s) present?"	384	1	0.16	378	1	0.97
	Q5: Have you heard about comparative genomics?	384	1	0.26	386	1	0.09
	Q6.1: Databases are free access resources.	137	1	0.10	327	1	0.16
	Q6.2: All citizens have access to the main genomic databases.	237	1	0.02*	341	1	0.02*
	Q6.3: All bioinformatics tools require programming skills.	316	1	0.95	368	1	0.28
	Q6.4: Bioinformatics tools are essential to molecular biology studies.	288	1	0.48	340	1	0.13

n – number of participants; df – degrees of freedom; Chi-Square test for a 95% confidence interval. (*) indicates significant differences between groups in each assessment moment.

Supplementary table III-3. Students knowledge, interest and attitudes towards bioinformatics integration at the science curricula – descriptive data.

		Control Group				Experimental Group			
		Pre-test		*Post-test*		*Pre-test*		*Post-test*	
		N	Mean	Std. Deviation	N	Mean	Std. Deviation	N	Mean	Std. Deviation	N	Mean	Std. Deviati
How important do you think bioinformatics is to ... (Q3)	... identify genes. (Q3.1)	92	4.35	0.78	93	4.83	0.41	276	4.45	0.68	291	4.74	0.52
	... store genomic data. (Q3.2)	92	4.36	0.82	94	4.51	0.58	273	4.37	0.82	291	4.53	0.71
	... study the evolutionary relations between organisms. (Q3.3)	80	3.81	0.83	93	4.01	0.96	239	3.74	1.09	276	4.05	1.01
Rate the importance of the following practices (Q9)	Practical work using digital tools in the classroom. (Q9.1)	95	4.42	0.66	95	4.54	0.68	286	4.46	0.65	288	4.51	0.63
	Study of genomes and gene regulation in bacteria. (Q9.2)	78	4.00	1.04	89	4.16	0.90	218	4.00	0.97	267	4.00	0.95
	Study of phylogeny / evolution of bacteria. (Q9.3)	71	3.96	0.92	83	3.80	1.15	230	3.96	0.99	270	4.05	0.92
	Using bioinformatics tools in the class. (Q9.4)	45	3.84	0.85	85	3.81	0.96	191	3.92	0.98	266	3.80	0.97
Rate your agreement with the following statements (Q7)	Different taxonomic groups of bacteria have genes in common. (Q7.1)	69	3.87	0.86	84	4.21	0.84	191	3.82	1.03	263	4.25	0.79
	There are different initiation codons. (Q7.2)	93	3.08	1.82	95	2.88	1.79	286	2.34	1.76	285	2.58	1.81
	There is a specific genetic code for bacteria. (Q7.3)	74	2.68	1.47	83	2.71	1.37	223	2.53	1.48	249	2.78	1.39
	All the bacterial genes are known. (Q7.4)	74	2.14	1.05	78	2.19	1.01	218	1.98	1.02	264	1.98	1.14
How often do you use the computer/technological devices... (Q8)	... for autonomous study outside the classroom. (Q8.1)	95	4.06	1.15	95	4.04	1.08	292	4.05	1.07	291	3.93	1.17
	... in the classroom to study. (Q8.2)	95	2.47	1.33	95	2.67	1.24	292	2.63	1.35	291	3.00	1.13
	... to access bioinformatics tools outside the classroom. (Q8.3)	95	1.41	0.82	95	1.38	0.81	287	1.56	0.99	291	1.77	1.06
	... to access bioinformatics tools in the classroom. (Q8.4)	95	1.55	0.99	95	1.55	0.90	288	1.71	1.06	291	2.61	1.05

N – Number of participants

Supplementary table III-4. Comparison between control and experimental group' knowledge, interest and attitudes towards bioinformatics.

		Pre-test				Post-test			
		t	*df*	*p*	|d|	*t*	*df*	*p*	|d|
Interest	*How important do you think bioinformatics is to ... (Q3)*								
	... identify genes. (Q3.1)	-1.16	140	0.25	0.20	1.65	195.1	0.10	0.24
	... store genomic data. (Q3.2)	-0.11	363	0.91	0.01	-0.27	383	0.91	0.03
	... study the evolutionary relations between organisms. (Q3.3)	0.62	177.8	0.54	0.09	-0.34	367	0.74	0.04
	Rate the importance of the following practices (Q9)								
	Practical work using digital tools in the classroom. (Q9.1)	0.48	379	0.63	0.05	0.39	381	0.70	0.04
	Study of genomes and gene regulation in bacteria. (Q9.2)	0.00	294	1.00	0.00	1.40	354	0.16	0.15
	Study of phylogeny / evolution of bacteria. (Q9.3)	0.01	299	0.99	0.00	-1.87	116.3	0.07	0.35
	Using bioinformatics tools in the class. (Q9.4)	-0.45	234	0.65	0.06	0.06	349	0.95	0.01
Knowledge	*Rate your agreement with the following statements (Q7)*								
	Different taxonomic groups of bacteria have genes in common. (Q7.1)	-0.42	144.04	0.68	0.07	-0.37	345	0.72	0.04
	There are different initiation codons. (Q7.2)	3.46	377	<0.01*	0.36	1.44	378	0.15	0.15
	There is a specific genetic code for bacteria. (Q7.3)	0.72	295	0.48	0.08	-0.37	330	0.71	0.04
	All the bacterial genes are known. (Q7.4)	1.11	290	0.27	0.13	1.51	340	0.13	0.16
Attitudes	*How often do you use the computer/technological devices... (Q8)*								
	... for autonomous study outside the classroom. (Q8.1)	0.09	385	0.93	0.01	0.84	384	0.40	0.09
	... in the classroom to study. (Q8.2)	-0.99	385	0.33	0.10	-2.30	148.5	0.03*	0.38
	... to access bioinformatics tools outside the classroom. (Q8.3)	-1.47	192.4	0.14	0.21	-3.79	205.3	<0.01*	0.21
	... to access bioinformatics tools in the classroom. (Q8.4)	-1.38	171.2	0.17	0.21	-9.62	184.9	<0.01*	1.42

t – independent samples t-test for a 95% confidence interval (*p*). *df* – degrees of freedom. *d* – Cohen's *d* measure of effect size. (*) indicates significant differences between groups in each assessment moment.

Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques

Abstract. *Predictive microbiology is a major opportunity to introduce youngsters to science of food preservation and food safety. In this study, we report a hands-on activity using ComBase to engage high school science students (15-16 years-old) in discussing real-world issues using scientific evidence. Moreover, it is expected to enhance participants' perception about the importance of bioinformatics and computational biology. This dry lab proposal revealed a positive impact on participants' knowledge on food microbiology and food preservation techniques, while promoting youngsters' citizenship education. Insights into participants' perceptions about the importance of computer applications to biological research were also gathered.*

Keywords. Bacteria, Bioinformatics, Food Preservation, Predictive Microbiology

1. Introduction

Food preservation involves the design of procedures to maintain the organoleptic properties and safety of perishable foods as long as possible (Rahman, 2007). Food safety and storage has been an ancient human activity. In fact, the early food preservation techniques such as drying, smoking, salting, chilling and heating are millenary, and have preceded the optimization of other methods technically more demanding (Zeuthen & Bogh-Sorensen, 2003).

In the nineteenth century, Pasteur' studies strongly contributed to understand the microbial nature of food spoilage, and therefore were decisive to improve food preservation methods (Zeuthen & Bogh-Sorensen, 2003). The scientific advances made established the foundations of modern food preservation methods, frequently characterized by technologically complex solutions. Altogether, these techniques are essential to increase food preservation, to allow the global trade of food goods and to raise consumers' confidence. Nowadays, food preservation is a highly interdisciplinary field of applied science with and enormous societal impact (Rahman, 2007).

Beyond the importance of food preservation strict sense, food safety is a major concern of stakeholders of agri-food production and supply chain. Not surprisingly, currently many studies are focused on food borne pathogens which might lead to serious illnesses or even be lethal. *Listeria monocytogenes* and pathogenic strains of *Escherichia coli* are particularly important food borne pathogens, commonly associated to several outbreaks. When these food poisoners bacteria are present in food products, they may pose a high risk for public health.

In Portugal, an outbreak of *Listeria* probably associated with cheese, occurred between March 2009 and February 2012. The outbreak resulted in 30 infected people and 11 deaths (Magalhães et al., 2015). In 2018, Austria, Denmark, Finland, Sweden and the United Kingdom have also reported 32 cases, with 6 deaths, caused by a *Listeria* outbreak probably due to frozen corn that was produced in Hungary and packaged in Poland (Food Safety News, 2018).

Regarding *E.coli*, in March 2017, a multi-state outbreak in the United States was studied. The possible cause for this incident was a nut-free substitute for peanut butter. 32 people were infected, being 12 of them hospitalized. 81% of ill people were younger than 18 years-old (Centers for Disease Control and Prevention, 2017). In 2011, the consumption of sprouted seeds and beans has been implicated in *E.coli* outbreaks in France and Germany (World Health Organization, 2011).

The mentioned examples highlight the importance of studying bacteria and the dynamics of their growth in order to avoid severe human health problems due to outbreaks. In this regard, modelling the growth or inactivation of bacteria in foodstuffs in response to extrinsic factors (e.g. temperature, oxygen concentration) and intrinsic factors (e.g. pH, NaCl concentration) is the objective of predictive microbiology. This research area, which integrates microbiology, mathematics, and computer science, is nowadays instrumental to food industry (Fakruddin et al., 2011; Perez-Rodriguez et al., 2013).

Predictive microbiology takes advantage of bacteria growth databases to validate the robustness of predictive models, contributing to a more accurate assessment of microbial risk (McKellar & Lu, 2003). *Pathogen Modelling Program* (Buchanan, 1993) and *ComBase* (Baranyi & Tamplin, 2004) are two open access web-based applications which are frequently used in food industry by risk assessors and food microbiologists. These applications allow users to define factors inputs and observe model outputs in informative graphics (McKellar & Lu, 2003).

The design of educational activities to promote the use of these computer-based platforms in the classroom is a major opportunity to introduce youngsters to science of food preservation and food safety. By stressing the importance of food safety and the need of actions in order to prevent foodborne illnesses, recent studies highlight the importance of more educational actions dedicated to food production practices taking into account the consumers' lack of knowledge on food safety (Badrie et al., 2006; Ergönül, 2013; Godwin et al., 2005; Sharif & Al-Malki, 2010). It is reported that TV and radio programs are the main vehicles to share information with consumers about food safety issues. However, educational initiatives promoted by governmental entities are crucial as they are seen by the consumers as more trustable (Bruhn & Schutez, 1999; Ergönül, 2013; Röhr et al., 2005). Moreover, there are

reported benefits of food safety and food preservation educational programs in the consumer's acceptance of new food preservation techniques, such as food irradiation (Angelillo et al., 2000; Kennedy et al., 2005; Pohlman, 1994).

According to the importance of this issue, food preservation techniques and its relationship with real-word issues such as food production, food trade and public health, are addressed in middle and high school level according to the Portuguese curricula (Bonito et al., 2013; Mendes et al., 2004; Silva et al., 2001), but also according to the Next Generation Science Standards for engineering, technology, and applications of science (National Research Council, 2013a, 2013d).

1.1. The Activity

In this work we propose three activities identified bellow using *ComBase*. These exercises were designed to estimate the growth/inactivation behavior of two well-known foodborne bacteria, *Listeria monocytogenes* and *Escherichia coli* aimed to foster knowledge about food safety and preservation techniques:

1st *ComBase* activity - Evaluate *Escherichia coli* growth in food matrices under different conditions of temperature, pH and NaCl.

2nd *ComBase* activity - Compare *E. coli* and *L. monocytogenes* growth in the same food matrix.

3rd *ComBase* activity - Determine the minimal lethal temperature required to inactivate *E. coli* and *L. monocytogenes*.

These activities focus on determining how extrinsic (temperature) and intrinsic factors (pH and NaCl) affect bacterial growth on food matrices, and on how these factors can be controlled to improve food preservation and increase food safety.

By accessing computational resources routinely used by food industry professionals, as *ComBase*, the ambition of these activities is to engage youngsters in discussing real-world issues using scientific evidence and foster their scientific literacy regarding food safety, food borne pathogens, public health and food preservation techniques. In addition, it is expected that these *in silico* exercises may enhance participants' perception about the importance of mathematics and computer science to tackle biology related questions, which one might generally designate as bioinformatics and computational biology.

2. Methods

2.1. Participants

The three *ComBase* activities within the scope of this study were implemented in July 2017 with 26 participants (high school science students with 15 ± 0,98 years old). At the time the activities were carried out, the participants (8 males and 18 females) had just finished their scholar year: the 9th (*n*=20), 10th (*n*=3) and 11th (*n*=3) grades.

These activities were implemented during a non-formal science education summer project designated "*Bacteria, Antibiotics and Resistance: let´s find out the links?*" within the scope of Junior University (*UJr*). This initiative of Porto University joins each year a few thousands of participants between 10 and 18 years old, split through dozens of summer-courses crossing thematically all scientific fields and planned according their school levels and background knowledge. The courses are hosted at different Faculties during four weeks of July and two monitors, usually undergraduates or graduated students, are allocated to each summer-course of 16 participants (Universidade do Porto, 2017). *UJr's* organizing committee and a board of Faculty members where the activities are performed, analyze all the activities proposals and decide regarding their approval.

To join this summer project the participants have to register using an online form available at the *UJr* site. The first sixteen candidates to complete the registration are automatically selected after the payment of a registration fee of 80 euros (*per* week).

In what concerns this study, in the beginning of the activity, each participant was invited to take part in this study after being introduced to its objectives. Students were informed that the participation in the study was not mandatory, so they could perform the activities without taking part in the study. All the data collected were processed and analyzed anonymously.

2.2. Data Collection

The assessment of activities impact on participants' knowledge about food microbiology, food safety and preservation, as well as their interest towards the use of computational resources to address life sciences questions, was carried out by a mixed method approach with a pre/post-test design to obtain a more robust appraisal of the results (Black, 1999; Punch, 2009). The pre/post questionnaire included closed questions (dichotomous and Likert-type scale format) and open-ended questions (Figure III - 7).

Questionnaire

Q1: Have you heard about Bioinformatics?
Yes __ No __

Q2: Classify Bioinformatics according the following goals *(1 – Not important to; 5 – Very important to)*:
a) Gene identification
b) Storage of genomic data
c) Observation of cell structures
d) Study of tridimensional structure of proteins
e) Study of phylogenetic relationships between organisms
f) Study bacteria growth

Q3: Classify as *True or False*:
a) The access to bioinformatics platforms is always paid
b) Bioinformatics tools are freely available to the public
c) Programming skills are needed to use bioinformatics tools
d) Bioinformatics resources include essential tools to optimize food preservation techniques

Q4: Which type of interaction between bacteria and food goods do you know?

Q5: Give some examples of food preservation techniques used to inhibit bacteria growth.

Q6: Which are the main causes of food spoilage that you know?

Q7: Rate your agreement regarding the following sentences (*1 – I totally disagree to 5 – I totally agree*):
a) The growth of microorganisms accelerates food spoilage
b) Food goods are spoiled by the same microorganisms
c) Food preservation techniques ensure sterilized food
d) There are food preservation techniques to preserve food goods

Figure III-7. Questionnaire used in this study.

The survey was complemented by naturalistic observations, which were aimed to identify misconceptions and evaluate participants' interaction and engagement. All descriptive and inferential statistical analyses were carried out using IBM Statistical Package for the Social

Sciences (SPSS) version 24 (Pallant, 2007). The content analysis of qualitative data included an inventory of concepts referred in the pre- and post-test (Weber, 1990).

2.3. Data Analyses

Questionnaire data were recorded, codified and categorized. All descriptive and inferential statistical analyses were carried out using IBMS' Statistical Package for the Social Sciences (SPSS) version 24.

Given the sample size and the nature of the research, i.e., quasi-experimental study, non-parametric tests were used: Wilcoxon test (paired samples) and McNemar test (dichotomous variables). In addition, an "*effect size test*" was carried out through the calculation of correlation coefficient (*r*) (available at: https://www.polyu.edu.hk/mm/effectsizefaqs/calculator/calculator.html).

In what concerns the qualitative data, a content analysis of the responses provided to each question was carried out. In the present study, this analysis included an inventory of concepts referred to in the pre-test and post-test.

It is important to mention that after deciding to analyze the data through non-parametric tests, complementary analyzes were also performed with parametric tests for critical analysis and validity.

3. Results and Discussion

3.1. Observation Records

During the implementation of the activity, participants worked in pairs or alone, according to their own choice. All the activities were supervised by two monitors, who were last year undergraduates in biochemistry and hired as monitors for this summer project by *UJr*.

While performing the exercises, youngsters discussed the outputs retrieved by the experiments and helped each other when some doubts came up. It was interesting to observe that participants were enthusiastic about the information they were finding in *Combase*. Curiously, participants also discussed between each other questions such as: "*Are the optimal growth conditions similar for all bacteria species*?" and then they autonomously tried to find an answer by searching for information at *ComBase*.

Monitors guided the participants through the workflow of the activities and encouraged participants to discuss, for example, about intrinsic and extrinsic factors that influence bacteria growth, highlighting the connection between the exercises performed and industry procedures of food preservation and safety. During this discussion an important comment was made by a participant: "*Ok, we can manipulate the conditions to avoid or to inhibit bacteria growth in*

food, but what about the organoleptic properties of the food goods?". The discussion lead to the conclusion that *"We can't just add salt or change pH adding vinegar to preserve the food good because the taste will not be the same! We have to find an alternative…*"

Participants' were challenged to present the results to their labmates, which has stimulated their scientific argumentation skills.

3.2. Questionnaire Analysis

This study showed that participants recognize metabolic activities (fermentation; aerobic respiration) as the main processes by which foodborne microorganisms mobilize the required nutrients and energy for growth (Q4) (Figure III - 8). An exact McNemar's test determined that there were no statistically significant differences in the proportion of this notion between pre- and post-test (p=0.45).

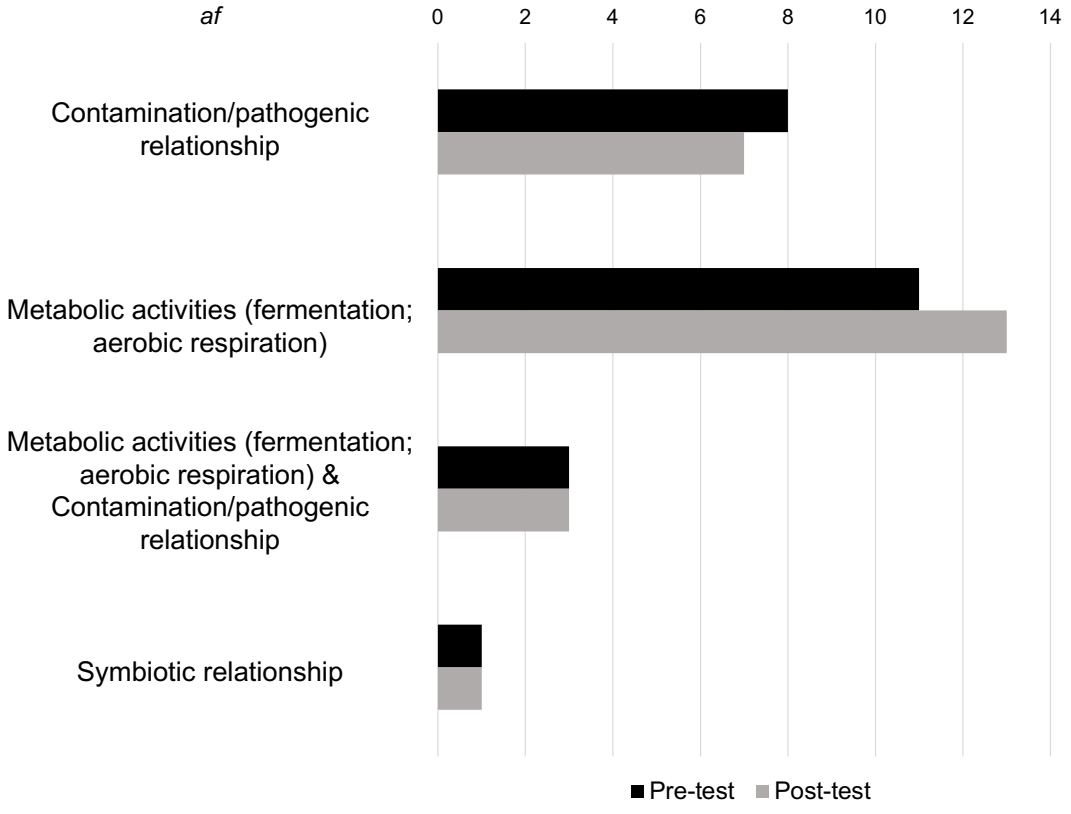


Figure III-8. Examples of participants' answers to the question Q4: "Which type of interaction between bacteria and food goods do you know?"; af - absolute frequency.

This association between foodstuffs and bacteria seems to align with participants school knowledge, since these themes are extensively addressed in the curricular contents of science classes at different grades (Mendes et al., 2001; National Research Council, 2013b, 2013c).

Regardless this evidence, participants acknowledge chemical oxidation as the main cause of food spoilage, in the pre-test (Q6) (Figure III - 9).

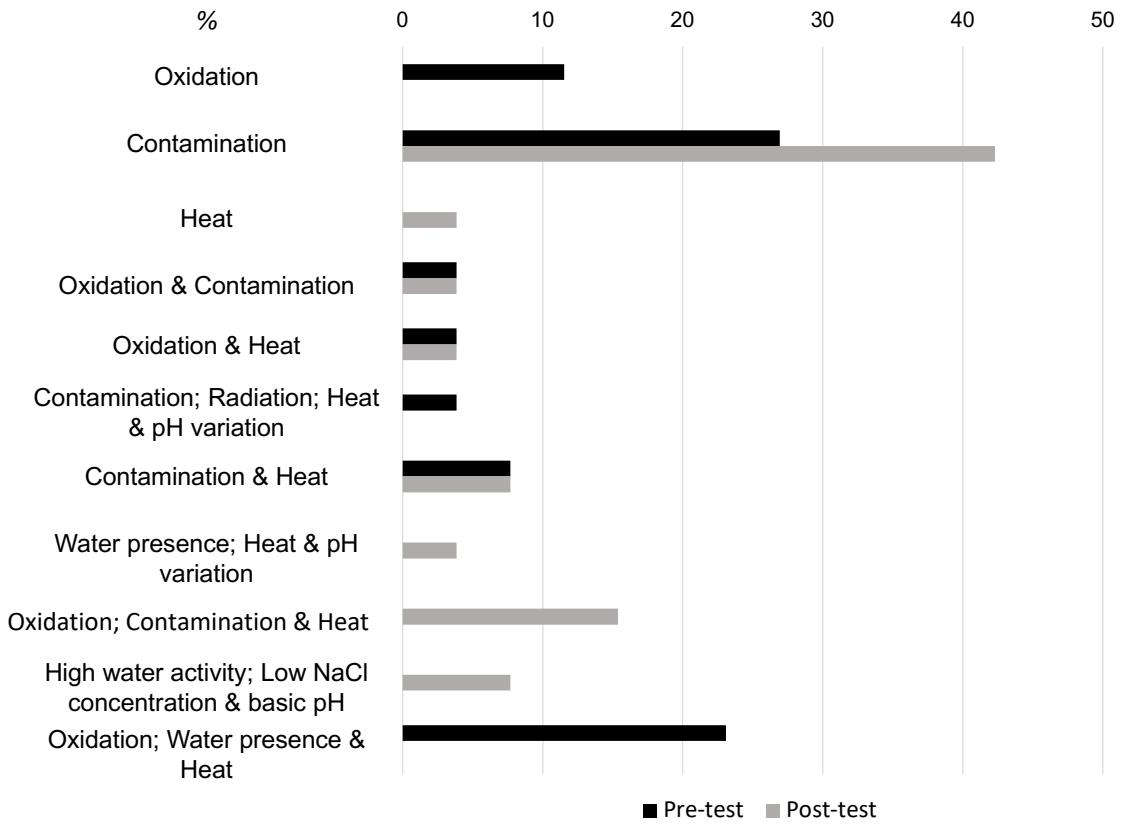


Figure III-9. Examples of answers to the question Q6: "Which are the main causes of food spoilage that you know?".

An exact McNemar's test determined that there was a statistically significant difference in the proportion of the notion "*oxidation*" between pre- and post-test (p=0.04). In the post-test, participants mentioned more frequently notions related with both bacteria and fungi contamination. However, there were no statistically significant differences in the proportion of the notion "*contamination*" between pre- and post-test (p=0.29). According to Trexler and Roeder (Trexler & Roeder, 2006), students can refer "*air*" and "*humidity*" as causes of food spoilage because they did not have the depth of vocabulary needed. In this scope, these terms were interpreted as being particularly suitable matrices for dissemination of "*bacteria*" or "*germs*", the true responsible for food spoilage. In fact, in our study, before the activity, students tended to dismiss explicitly microbial contamination by mentioning environmental conditions as the main causes for food spoilage, giving as an example chemical oxidation. After the activities, participants were more insightful in indicating microbial contamination as a significant cause of food spoilage.

After the *ComBase* hands-on activities, participants agree that "*The growth of microorganisms accelerates food spoilage*" (Q7 - a) and acknowledged the importance of "*Food processing techniques to preserve food goods*" (Q7 - d) (Figure III - 10). However,

differences between pre- and post-test are not statistically significant (Q7 – a: Z = -1.46, p=0.14, r = - 0.29; Q7 – d: Z = -1.9, p=0.06, r = - 0.39).

Although there are no statistically significant differences, the changes in the answer's average values, suggest that participants awareness about food spoilage and food borne illnesses by specific bacteria such as *E. coli* and *L. monocytogenes*, has been improved. This finding meets scientific evidence about food illness, due to the risk of infection by the proliferation of these agents in the human body after consumption of contaminated food (Tent, 1999). On the other hand, participants tend to disagree that "*Foods are spoiled by the same microorganisms*" (Q7 – b) (Figure III - 9). A Wilcoxon signed-rank test showed that there were not statistically significant differences in the frequency of this notion between pre- and post-test (Z = -0.78, p=0.44, r = - 0.16). This notion aligns well with data showing that food matrices can be differently contaminated by microorganisms, depending on the chemi-physical properties of the food good. Furthermore, participants tend to accept the scientific evidence highlighting the wide diversity of microorganisms that, under some conditions, can contribute to food contamination and spoilage (Gram et al., 2002).

It is interesting to note that participants seem to recognize the difference between food preservation and sterilization, once they tend to disagree with the sentence "*Food preservation techniques do not ensure sterilized food"* (Q7 – c). However, the difference between pre- and post-test is not statically significant (Z = -1.26, p=0.21, r = - 0.26).

Regarding food preservation techniques (Q5), it is noteworthy that in the pre-test, participants pointed out low temperature processes (e.g. "*cooling*", "*freezing*") as the main food preservation techniques (Figure III - 11), implicitly dismissing the importance of other food preservation methods such as the use of chemical preservatives and modified atmosphere packaging (Figure III - 11). These outcomes were expected as low temperature processes are widely known and frequently used as a domestic food preservation practice. However, studies have emphasized consumers' bad practices regarding temperature control of domestic fridges and handling leftovers, highlighting the need to promote educational approaches dedicated to food safety practices (Bruhn & Schutez, 1999; Redmond & Griffith, 2003). Although no specific assessment was made, after the exercises the participants seemed to recognize the importance to use correctly the fridge, respect the expiration date of perishable goods, and carefully handle leftovers. Furthermore, in the post-test, although participants kept acknowledging low temperatures as a key extrinsic factor used to inhibit bacterial growth in foodstuffs, the frequency of the notions related with the "*use of bacteria growth inhibitors*" and "*the preservation* of *food by a modified atmosphere*" increase significantly from pre- to post-test, according to McNemar's test (p=0.01, p=0.02; respectively) (Figure III - 11).

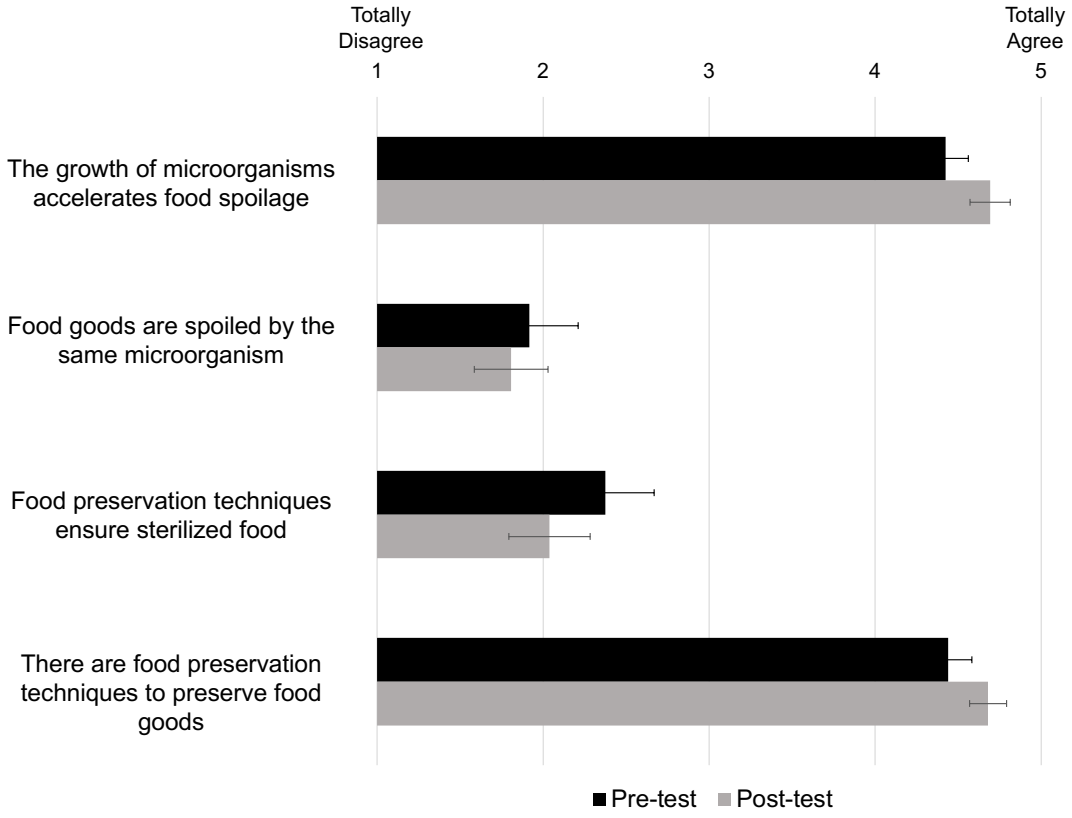


Figure III-10. Students ranked their agreement with the listed notions regarding food spoilage and preservation techniques (Q7).

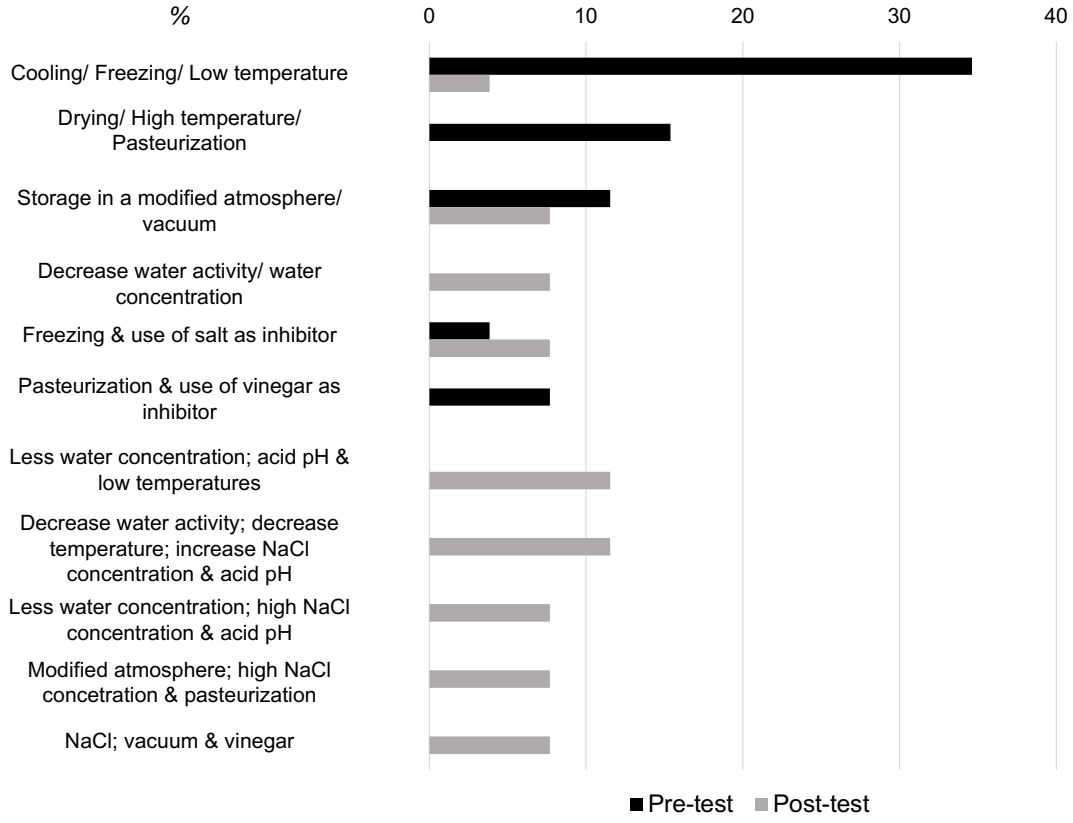


Figure III-11. Examples of answers to the question Q5: "Give some examples of food preservation techniques to inhibit bacteria growth".

These results suggest that participants become more aware of the chemical bacteria growth inhibitors available, including natural compounds as oil and plant extracts which benefits are currently being highlighted (Ahn et al., 2007; Hammer et al., 1999; Negi, 2012). In this regard, the activity described by Fonseca & Tavares (Fonseca & Tavares, 2011) is a good example to combine up-to-date scientific research issues with hands-on educational practices.

Taking into account that the proposed activities were centered on computer resources using *ComBase* to understand the impact of food preservation techniques on microbial growth, it was important to assess participants' perceptions and interest about computational biology tools and bioinformatics.

The results showed that most participants had never heard about bioinformatics (Q1) (Figure III – 12). Although there were no statistically significant differences between pre- and post-test, after the activities the participants could highlight the potential of computational biology and bioinformatics to store data (Q2 – b) (Z = - 1.76, p = 0.08, r = - 0.39) and to study bacteria growth (Q2 – f) (Z = - 1.79, p = 0.07, r = - 0.36) (Figure III - 10). In fact, predictive microbiology software's use databases and mathematical models, both framed within the scientific fields of Bioinformatics and Biomathematics (McKellar & Lu, 2003).

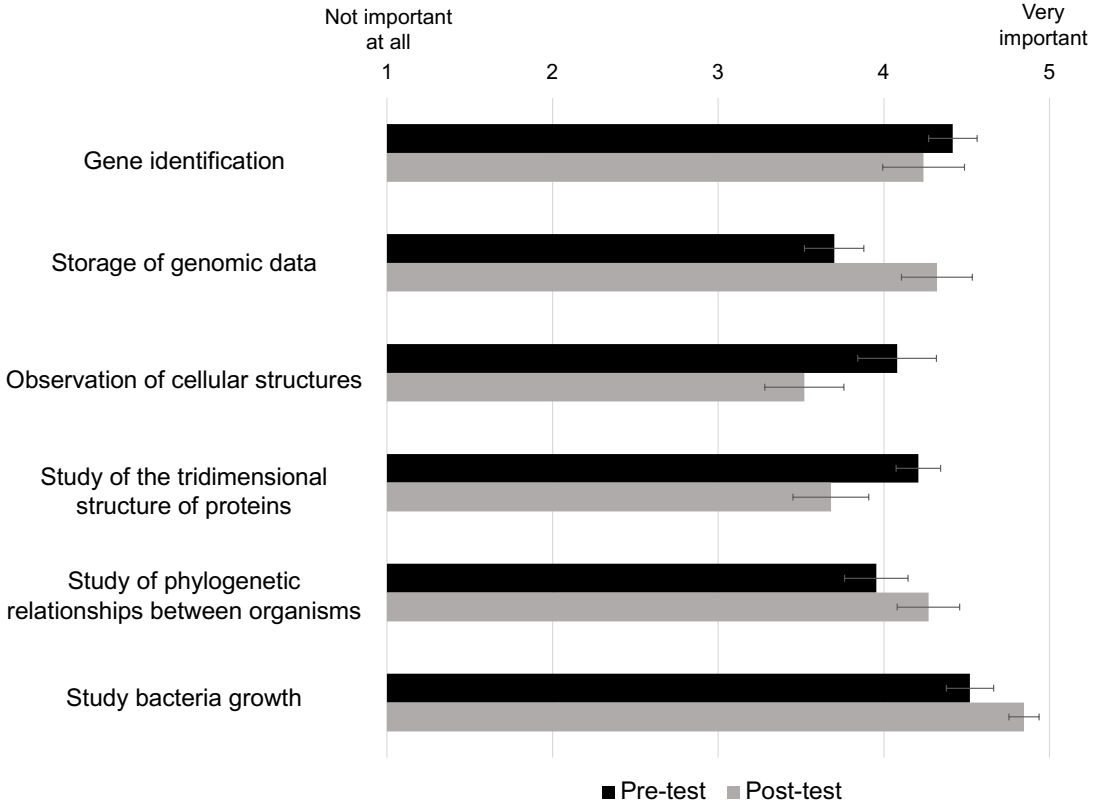


Figure III-12. Notions regarding bioinformatics applications that participants agree with (Q2).

Beyond the knowledge acquired regarding food spoilage and preservation, with these activities the participants were able to address the importance of mathematics modelling and computer sciences. When challenged with specific questions regarding these issues, participants were able to overcome misconceptions such as *"One cannot freely access bioinformatics tools"* (Q3 – b) or *"Programming skills are needed to use bioinformatics tools"* (Q3 – c). A McNemar's test determined that there was a statistically significant difference in the proportion of these two notions between pre- and post-test (p=0.04, p<0.01; respectively) (Figure III - 13). These results are particularly meaningful for elementary and high school teachers once the use of bioinformatics-based activities in the classroom is a unique opportunity to conciliate biological data with experimental *in silico* work, allowing students to get acquainted with open-access data, and with freely available computer applications running in personal computers (PCs) and web browsers with user-friendly interfaces. This is an education paradigm still poorly explored but essential to dower youngsters with reasoning skills to face the technological revolution that we live in (Livingstone, 2012). In the context of this summer project, learning from scientists or through scientific tools about food spoilage and contamination using online applications, was a great opportunity to increase their interest in emerging areas such as genomics and genomic information (Haga et al., 2013). Moreover, youngsters realize that biologists might easily take advantage of bioinformatics, although their background is different from a bioinformatics technician (Pevzner & Shamir, 2011). Nowadays, there is an increased awareness to develop data analysis software that provide bioinformatics functionalities to biologists without requiring prior knowledge on computer science (Mandoiu & Zelikovsky, 2008; Kumar & Dudley, 2007).

When addressing the importance of bioinformatics resources to food industry, students recognized that "*Bioinformatics resources include essential tools to optimize food preservation techniques*" (Q3 – d) (Figure III - 13). A McNemar test revealed no statically significant differences between pre and post-test (p=1.00). However, there is an indication of the positive impact of the implemented activities as reported in previous studies (Fernandes et al., 2014).

When enquired about the questionnaire, participants acknowledged its importance at both assessment moments (pre – and post-test) and considered that it was easy to understand (Figure III - 14).

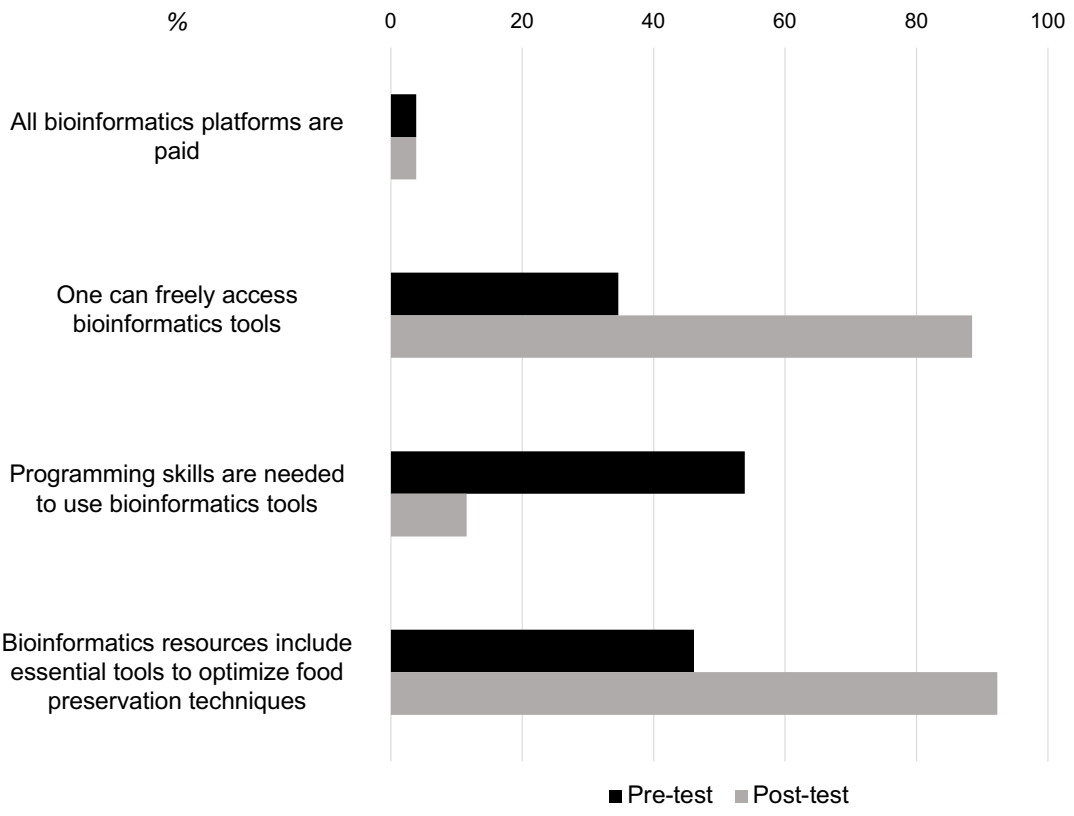


Figure III-13. Percentage of statements regarding bioinformatics that participants classified as true (Q3).

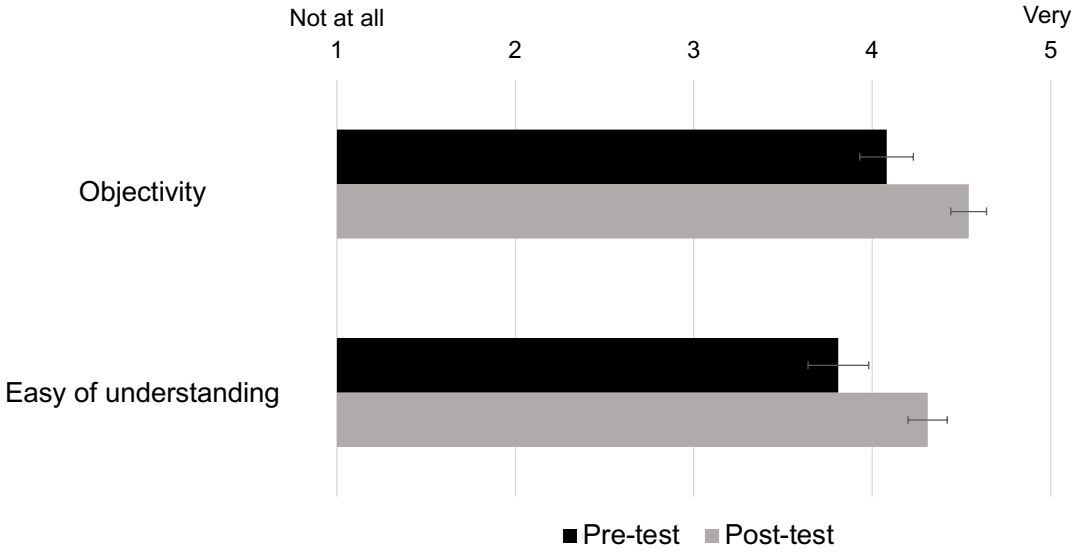


Figure III-14. Questionnaire assessment by students.

4. Conclusions and Future Perspectives

Overall, these dry lab activities revealed a positive impact on participants' knowledge about food microbiology and food preservation techniques, while promoting youngsters' citizenship education. In addition, the data gathered with this study provided insights into youngsters' perceptions about the importance of computer applications to biological research. These conclusions are in line with recent studies on the added value of *in silico* activities in the classroom, as a motivational driver to address other curricular themes, namely gene regulation and evolution (Martins et al., 2018), involving open-access bioinformatics resources and time-efficient exercises.

Acknowledgements

The authors are grateful to all the participants of the first edition of the activity "*Bacteria, Antibiotics and Resistance: let's find out the links?*" (*UJr* 2017) and to Leonor Martins for the fruitful comments made on the manuscript. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Ahn, J., Grün, I. U., & Mustapha, A. (2007). Effects of plant extracts on microbial growth, color change, and lipid oxidation in cooked beef. *Food Microbiology*, *24*(1), 7–14. https://doi.org/10.1016/j.fm.2006.04.006

Angelillo, I. F., Viggiani, N. M. A., Rizzo, L., & Bianco, A. (2000). Food handlers and foodborne diseases: Knowledge, attitudes, and reported behavior in Italy. *Journal of Food Protection*, *63*(3), 381–385. https://doi.org/10.4315/0362-028X-63.3.381

Badrie, N., Gobin, A., Dookeran, S., & Duncan, R. (2006). Consumer awareness and perception to food safety hazards in Trinidad, West Indies. *Food Control, 17*(5), 370-377. https://doi.org/10.1016/j.foodcont.2005.01.003

Baranyi, J., & Tamplin, M. L. (2004). ComBase: A common database on microbial responses to food environments. *Journal of Food Protection*, *67*(9), 1967–1971. https://doi.org/10.4315/0362-028X-67.9.1967

Black, T. (1999). *Doing Quantitative Research in the Social Sciences*. SAGE Publications Ltd. https://uk.sagepub.com/en-gb/eur/doing-quantitative-research-in-the-social-sciences/book205869

Bonito, J., Morgado, M., Silva, M., Figueira, D., Serrano, M., Mesquita, J., & Rebelo, H. (2013). *Metas Curriculares Ensino Básico - Ciências Naturais 5º,6º,7º e 8ºanos*.

Bruhn, C., & Schutez, H. (1999). Consumer Food Safety Knowledge and Practices. *Journal of Food Safety*, *19*(1), 73–87. https://doi.org/10.1111/j.1745-4565.1999.tb00235.x

Buchanan, R. L. (1993). Developing and distributing user-friendly application software. *Journal of Industrial Microbiology*, *12*(3–5), 251–255. https://doi.org/10.1007/BF01584198

Centers for Disease Control and Prevention. (2017, May 4). *Multistate Outbreak of Shiga toxin-producing Escherichia coli O157:H7 Infections Linked to I.M. Healthy Brand SoyNut Butter*. Centres for Disease Control and Prevention. https://www.cdc.gov/ecoli/2017/o157h7-03-17/index.html

Ergönül, B. (2013). Consumer awareness and perception to food safety: A consumer analysis. *Food Control*, *32*(2), 461–471. https://doi.org/10.1016/j.foodcont.2013.01.018

Fakruddin, M., Mazumder, R. M., Shahnewaj, K., & Mannan, B. (2011). Predictive microbiology: Modeling microbial responses in food. *Ceylon Journal of Science (Bio. Sci.)*, *40*(2), 121–131.

Fernandes, E., Dias, C., Fonseca, M. J., & Tavares, F. (2014). Understanding Growth and Thermal Inactivation of Foodborne Bacteria Using the Pathogen Modelling Program (PMP). In M. Costa, P. Pombo, & J. Dorrio (Eds.), *Hands-on Science: Science Education with and for Society* (pp. 207–210). Hands-on Science Network.

Fonseca, M. J., & Tavares, F. (2011). Natural Antibiotics: A Hands-on Activity on Garlic's Antibiotic Properties. *The American Biology Teacher*, *73*(6), 342–346. https://doi.org/10.1525/abt.2011.73.6.7

Food Safety News. (2018, March 23). *E. coli O157:H7 findings trigger multiple beef recalls in Canada | Food Safety News*. https://www.foodsafetynews.com/2018/03/e-coli-o157h7-findings-trigger-multiple-beef-recalls-in-canada/

Godwin, S., Coppings, R., Speller-Henderson, L., & Pearson, L. (2005). Study Finds Consumer Food Safety Knowledge Lacking. *Journal of Family and Consumer Sciences*, *97*(2), 40–44. https://eric.ed.gov/?id=EJ720620

Gram, L., Ravn, L., Rasch, M., Bruhn, J. B., Christensen, A. B., & Givskov, M. (2002). Food spoilage - Interactions between food spoilage bacteria. *International Journal of Food Microbiology*, *78*(1–2), 79–97. https://doi.org/10.1016/S0168-1605(02)00233-7

Haga, S. B., Rosanbalm, K. D., Boles, L., Tindall, G. M., Livingston, T. M., & O'Daniel, J. M. (2013). Promoting public awareness and engagement in genome sciences. *Journal of Genetic Counseling*, *22*(4), 508–516. https://doi.org/10.1007/s10897-013-9577-3

Hammer, K. A., Carson, C. F., & Riley, T. V. (1999). Antimicrobial activity of essential oils

and other plant extracts. *Journal of Applied Microbiology*, *86*(6), 985–990. https://doi.org/10.1046/j.1365-2672.1999.00780.x

Kennedy, J., Jackson, V., Blair, I. S., McDowell, D. A., Cowan, C., & Bolton, D. J. (2005). Food safety knowledge of consumers and the microbiological and temperature status of their refrigerators. *Journal of Food Protection*, *68*(7), 1421–1430. https://doi.org/10.4315/0362-028X-68.7.1421

Kumar, S., & Dudley, J. (2007). Bioinformatics software for biologists in the genomics era. *Bioinformatics*, *23*(14), 1713–1717. https://doi.org/10.1093/bioinformatics/btm239

Livingstone, S. (2012). Critical reflections on the benefits of ICT in education. *Oxford Review of Education, 38*(1), 9-24. https://doi.org/10.1080/03054985.2011.577938

Magalhães, R., Almeida, G., Ferreira, V., Santos, I., Silva, J., Mendes, M. M., Pita, J., Mariano, G., Mâncio, I., Sousa, M. M., Farber, J., Pagotto, F., & Teixeira, P. (2015). Cheese-related listeriosis outbreak, Portugal, march 2009 to february 2012. *Eurosurveillance*, *20*(17), 21104. https://doi.org/10.2807/1560-7917.ES2015.20.17.21104

Mandoiu,I., & Zelikovsky, A. (2008). *Bioinformatics Algorithms: Techniques and Applications*. John Wiley & Sons.

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

McKellar, R., & Lu, X. (2003). *Modeling Microbial Responses in Food*. CRC Press.

Mendes, A., Rebelo, D., & Pinheiro, E. (2004). *Biologia 12º ano - Curso Científico Humanístico de Ciências e Tecnologias*. Ministério da Educação.

National Research Council. (2013a). *HS-ETS1 Engineering Design | Next Generation Science Standards*. https://www.nextgenscience.org/dci-arrangement/hs-ets1-engineering-design

National Research Council. (2013b). *HS-LS1 From Molecules to Organisms: Structures and Processes | Next Generation Science Standards*. https://www.nextgenscience.org/dci-arrangement/hs-ls1-molecules-organisms-structures-and-processes

National Research Council. (2013c). *HS-LS2 Ecosystems: Interactions, Energy, and Dynamics | Next Generation Science Standards*. https://www.nextgenscience.org/dci-arrangement/hs-ls2-ecosystems-interactions-energy-and-dynamics

National Research Council. (2013d). *MS-ETS1 Engineering Design | Next Generation Science Standards*. https://www.nextgenscience.org/dci-arrangement/ms-ets1-engineering-

design

Negi, P. S. (2012). Plant extracts for the control of bacterial growth: Efficacy, stability and safety issues for food application. *International Journal of Food Microbiology*, *156*(1), 7–17. https://doi.org/10.1016/j.ijfoodmicro.2012.03.006

Pallant, J. (2007). *SPSS - Survival Guide to Data Analysis using SPSS for Windows*. Open University Press/McGraw-Hill.

Perez-Rodriguez, F., Valero, A., Pérez-Rodríguez, F., & Valero, A. (2013). Predictive Microbiology in Foods. In *Predictive Microbiology in Foods* (pp. 1–10). Springer New York. https://doi.org/10.1007/978-1-4614-5520-2_1

Pevzner,P., & Shamir, R. (2011). *Bioinformatics for Biologists*. Cambridge University Press.

Pohlman, A., Wood, O., & Mason, A. (1994). Influence of audiovisuals and food samples on consumer acceptance of food irradiation. *Food Technology*, *48*(12), 46–48.

Punch, K. (2009). Introduction to research methods in education. In *Introduction to research methods in education*.

Rahman. (2007). *Handbook of Food Preservation* (2nd edition). CRC Press.

Redmond, E. C., & Griffith, C. J. (2003). Consumer food handling in the home: A review of food safety studies. *Journal of Food Protection*, *66*(1), 130–161. https://doi.org/10.4315/0362-028X-66.1.130

Röhr, A., Lüddecke, K., Drusch, S., Müller, M. J., & Alvensleben, R. V. (2005). Food quality and safety-consumer perception and public health concern. *Food Control*, *16*(8), 649–655. https://doi.org/10.1016/j.foodcont.2004.06.001

Sharif, L., & Al-Malki, T. (2010). Knowledge, attitude and practice of Taif University students on food poisoning. *Food Control*, *21*, 55–60. https://doi.org/10.1016/j.foodcont.2009.03.015

Silva, C. P., Amador, F., Fernando, J., Baptista, P., Adérito, R., Colaboradores, V., Mendes, A., Rebelo, D., & Pinheiro, E. (2001). *Programa de Biologia e Geologia 10º ano*.

Tent, H. (1999). Research on food safety in the 21st century. *Food Control*, *10*(4–5), 239–241. https://doi.org/10.1016/s0956-7135(99)00025-0

Trexler, C. J., & Roeder, D. (2006). Using Qualitative Research Methods to Ascertain Elementary Students' Understandings of Food Safety. *Journal of Food Science Education*, *2*(2), 25–31. https://doi.org/10.1111/j.1541-4329.2003.tb00022.x

Universidade do Porto. (2017). *Universidade Júnior*. https://universidadejunior.up.pt/

Weber, R. (1990). *Basic Content Analysis*. SAGE Publications.

https://doi.org/10.4135/9781412983488

World Health Organisation. (2011). *Outbreaks of E.coli O104:H4 infection*. World Health Organization: Health Topics. http://www.euro.who.int/en/healthtopics/%0Demergencies/international-healthregulations/%0Dnews/news/2011/07/outbreaks%0D-of-e.-coli-o104h4-infection-whoeuropegives-%0Dpublic-health-advice

Zeuthen, P., & Bogh-Sorensen, L. (2003). *Food Preservation Techniques*. Woodhead Publishing Limited.

Chapter IV

Perceptions and Attitudes of Teachers towards Bioinformatics Education

Chapter IV includes the following publications:

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of Bioinformatics Tools to Elementary and Secondary School Curricula: a Training Course for Teachers. In *Congresso Internacional - O Tempo dos Professores.* (pp. 515-522). FPCEUP. ISBN: 978-989-8471-26-0

Martins, A., Lencastre, L., & Tavares, F. (2018). Integrating Bioinformatics in Elementary and Secondary Education: Teacher's Perceptions. In *3rd International Conference on Teacher Education (INCTE)*. (pp. 203-214) Instituto Politécnico de Bragança. ISBN: 978-972-745-241-5

Martins, A., Lencastre, L., & Tavares, F. (2020). Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers. *The Beauty and Pleasure of Understanding: Engaging with Contemporary Challenges Through Science Education*. (pp. 1712-1721). European Science Education Research Association. ISBN: 978-88-945874-0-1

Martins, A., Lencastre, L., & Tavares, F. (2020). Bioinformatics, a Befitting Tool for e-learning: Potential and Constraints according Teachers' Perceptions. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.). *Hands-on Science. Science Education. Discovering and understanding the wonders of Nature.* (pp. 97-105). Hands-on Science Network. ISBN: 978-84-8158-841-5

Part of the results from this chapter were presented at:

Martins, A., Lencastre, L., & Tavares, F. (2019). Bioinformatics in Secondary Education: From Wishful Thinking to Reality. Microbiotec 19. Coimbra, 5th-7th December 2019.

Adequacy of Bioinformatics Tools to Elementary and Secondary School Curricula: a Training Course for Teachers

Abstract. *Bioinformatics is the use of computational resources to categorize massive raw data in datasets and retrieve meaningful information. This relatively recent field of biological research highlights the need to adequate elementary and secondary curricula and educational resources to this new paradigm.*

In this context, is important to acknowledge the central role of teachers as instrumental educational agents to implement and foster learning through bioinformatics tools. To achieve this objective, it is urgent to implement specialized training courses for teachers, allowing them to explore confidently bioinformatics' tools with a considerable degree of demand. Furthermore, recent studies have shown that, after a first foray into bioinformatics activities, teachers tend to feel more motivated, confident and able to redesign the didactic-pedagogical goals.

Aiming to improve teachers' knowledge in bioinformatics and to foster innovative teacher practices, a training course for teachers entitled "Adequacy of Bioinformatics Tools to Elementary Education and Secondary Education", has been implemented at the Faculty of Sciences, UP. The main goal of this course, is to provide teachers with the knowledge to implement in their educational practices user-friendly and open access bioinformatics tools and resources, conveniently aligned with the ongoing curricular context. Regardless teachers' enthusiasm about bioinformatics, the main constraints identified by teachers were deficient school facilities regarding computers and internet access; English based-language of bioinformatics platforms; and teachers' poor background in the field. On the other hand, participant teachers highlighted that the training course contributed to increase their confidence to implement bioinformatics' activities in their classes, which they believe will have a beneficial impact on student's learning and encourage student's interest on STEM related issues.

Keywords. Bioinformatics, Elementary and Secondary Curricula, Professional Development, Teaching

1. Introduction

Nowadays computers play a primordial function in routine biological analysis, either by dealing with big datasets, and/or retrieving meaningful information. Bioinformatics is

a new paradigm of biological research, mainly boosted by the enormous technological developments on DNA sequencing brought by high-throughput Next Generation Sequencing (NGS) platforms. This relatively recent field of biological research call for the need to rethink biological education and adequate elementary and secondary curricula and educational resources to this new paradigm. Several studies have emphasized the beneficial impact of bioinformatics-based activities in students learning on topics such drug-resistance, evolution and gene expression (Amenkhienan & Smith, 2006; Newman et al., 2016; Taylor et al., 2014). Recent findings indicate that students involved in bioinformatics activities, apart from the conceptual knowledge, improve their computer skills and their ability to adapt to the digital era society (Machluf et al., 2017). Despite the gathered evidence, the poor teachers' training in basics bioinformatics is probably the major constraint to implement curricular framed bioinformatics exercises. It is worth to recall that bioinformatics-related topics were not comprised in pre-Bologna university curricula. With the transition to post-Bologna, although bioinformatics was introduced in the curricular structure of several scientific oriented master's degrees, it was left out of teachers' education Masters. In fact, biology teachers who are currently getting their academic degrees, have a poor training in bioinformatics, and feel unsure and misinformed to use bioinformatics tools in their teaching practices, as previously reported regarding biotechnology. To face this unfavorable landscape of teacher training, it is absolutely urgent to offer adequate pre-service and in-service teachers' training courses in bioinformatics, as well as provide an assortment of didactic-tools to assist teachers in the implementation of bioinformatics-based activities, but also by lighten their busy agenda (Marques et al., 2014; Shuster et al., 2016; Wood & Gebhardt, 2013). Recent studies have shown that, after a first foray into bioinformatics activities, teachers tend to feel more motivated, confident and able to redesign the didactic-pedagogical goals to be achieved (Machluf et al., 2017). Not surprisingly, international reputable biological research institutions, namely the European Molecular Biology Laboratory (EMBL) (European Molecular Biology Laboratory, 2017), and Netherlands Bioinformatics Centre (NBIC) (Netherlands Bioinformatics Centre, 2017), have provided on-line support to teachers willing to include bioinformatics in their teaching practices.

Acknowledging the role of educators as promoters of students' achievements and ambitioning to improve teachers' knowledge in bioinformatics and foster innovative teacher practices, a training course for teachers entitled "*Adequacy of Bioinformatics Tools to Elementary Education and Secondary Education*", has been implemented at the Faculty of Sciences, UP. This practical course for in-service biology teachers, includes a portfolio of bioinformatics activities selected and adapted to meet the Portuguese

national curricula. In addition to the training this course, a webpage (https://bioinformaticaaula.wixsite.com/bioinformatica-pt) was created to provide continuous support. The assessment of the first edition of this course, carried out in March 2017, allowed to gather data about teachers' perceptions concerning bioinformatics and its adequacy to the science teaching and learning context.

2. The Training Course and Participants

The main goal of the training course on "*Adequacy of Bioinformatics Tools to Elementary Education and Secondary Education*", which was accredited by the competent Portuguese body (register nº CCPFC/ACC-88413/16) and confer credits to participant teachers, is to provide teachers with the knowledge and competences to implement in their educational practices user-friendly and open access bioinformatics tools and resources, conveniently aligned with the ongoing curricular context. The training course included nineteen exercises, including step-by-step guidelines for teachers to feel more comfortable about the procedures (Form & Lewitter, 2011; Machluf & Yarden, 2013) (Figure IV - 1).

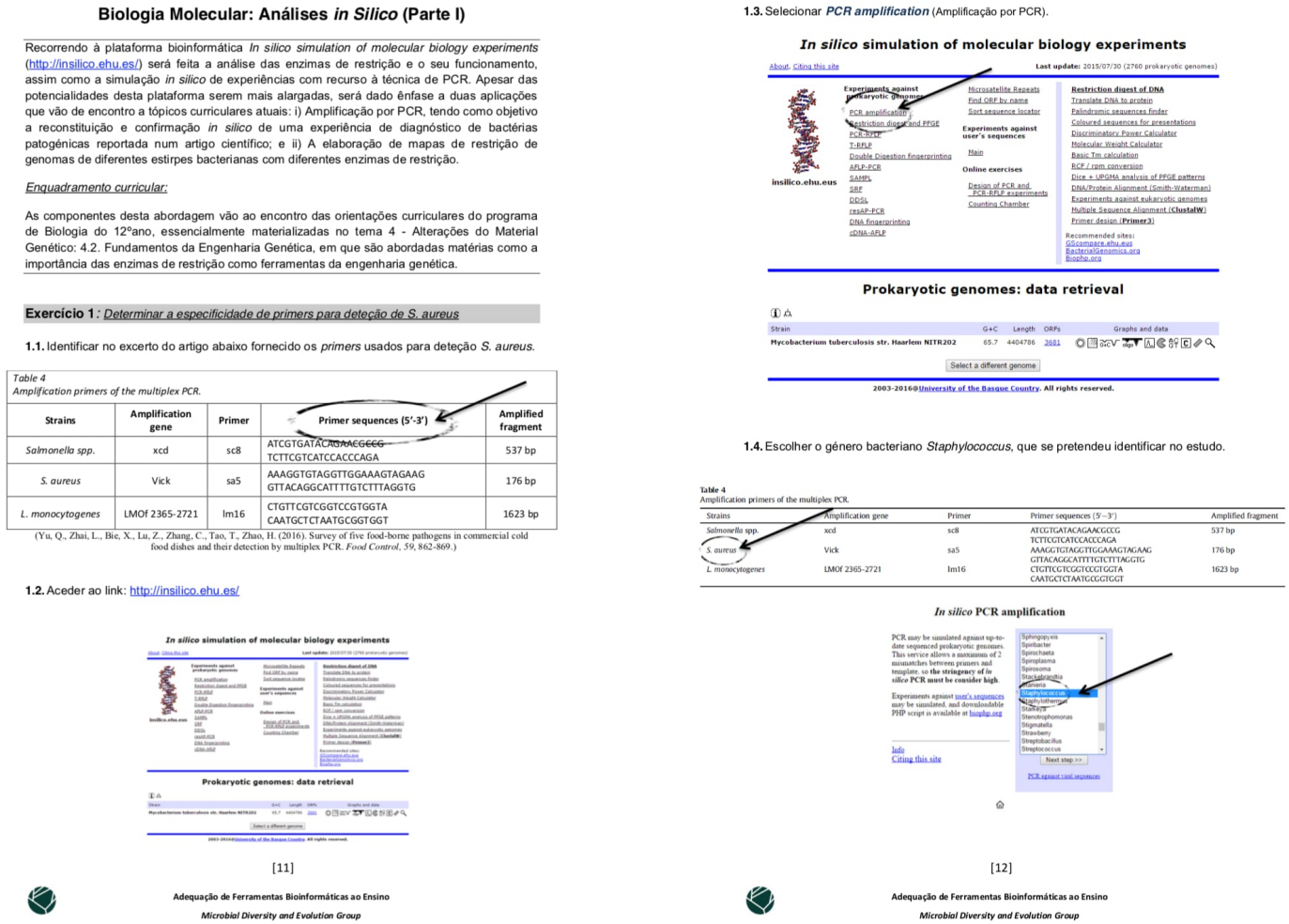


Figure IV-1. Example of the guidelines designed for the training course.

The course contents include five major topics, to facilitate an array of options to teachers practicing at different levels: 1) *In silico* analysis of molecular biology experiments; 2) Lac operon: gene regulation; 3) Importance of genes and mutations for evolution; 4) Exploring metabolic pathways at different domains of life; and 5) Bioinformatics impact on society: practical examples.

Thirteen in-service teachers (n=13) from twelve schools (10 public and 2 private) belonging to Porto metropolitan area, participated in the first edition of the training course "*Adequacy of Bioinformatics Tools to Elementary Education and Secondary Education*". The participants teaching experience was 22.92 ± 4.98 years and the majority of the participants (11 out of 13) were teaching at the secondary school level.

3. Methodology

To unveil participant teacher's perception about bioinformatics education and diagnose their acquired skills in basics bioinformatics, an interview-based methodology as well as a survey was designed and applied, covering a range of topics, namely: academic background and sense of proficiency; attitudes towards bioinformatics; and perceptions about the availability of bioinformatics resources. The survey was measured in a five-point Likert-type scale.

IBMS' SPSS was used to analyze the data.

4. Research Findings

The main objective of this teacher's training course was to grant the group of teachers with the basic tools required to foster learning based on suitable bioinformatics platforms. This intervention allowed gathering some evidences regarding teacher's perceptions and attitudes about bioinformatics.

The main constraints pointed out by the teachers to implement bioinformatics-based labs in the classroom were poor computer facilities and weak internet access in some schools. In addition, some teachers admit difficulties related with the English language of the bioinformatics platforms. Regardless these difficulties highlighted by the participant teachers, it was evident that the training course contributed to nurture their enthusiasm about bioinformatics. Furthermore, they feel that with convenient support, e.g. hardware, and more training opportunities, would increase decisively their confidence to implement bioinformatics' activities in their classes (Figure IV - 2). In fact, the survey analysis showed that teachers' interest in bioinformatics is high (M = 4.92, SD = 0.277) and that teachers' perception about their knowledge regarding bioinformatics increased during the training course (M before course = 2, SD = 0.816, M

after course = 3.62, SD = 0.961). When asked if "*The training courses available in the field of bioinformatics are still scarce*" it is clear that the participants agree with the statement (M = 4.46, SD = 0.967). Twelve out of the thirteen participants show interest in participating in training actions / courses promoted by research groups that routinely resort to bioinformatics tools.

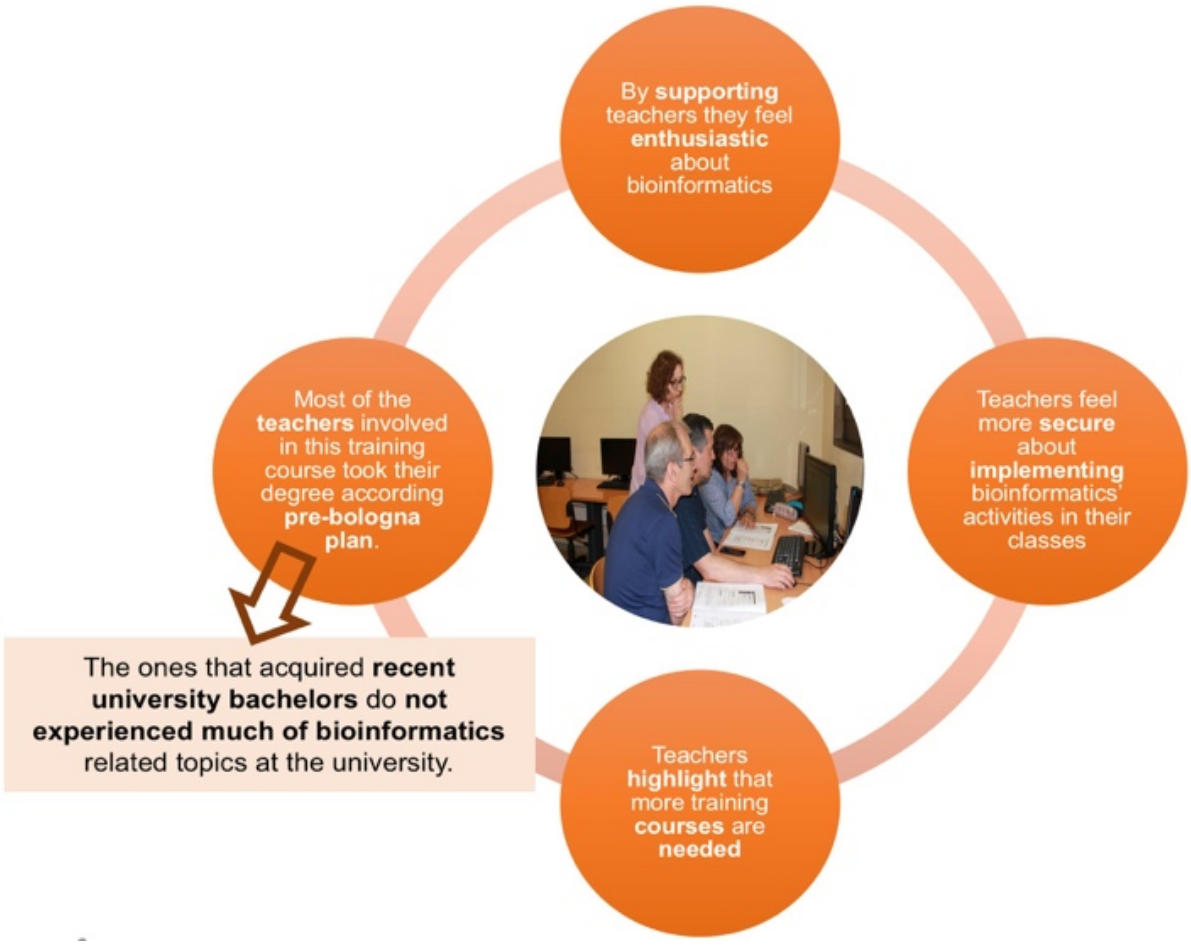


Figure IV-2. Teachers' Bioinformatics training course: main findings.

There is clear agreement among the teachers about the importance of bioinformatics for current research / scientific advances (M = 4.92, SD = 0.277). Adding to this, they highlighted the importance of using bioinformatics tools both in the elementary school level (M = 3.55, SD = 0.293) as well as in secondary school level (M = 4.54, SD = 0.66). Thus, it is undisputable that teachers recognize the importance of bioinformatics for biological research, but more importantly believe that bioinformatics might be useful in their teaching routines, even if they agree that "*Preparing a class using bioinformatics tools requires more time and resources than preparing other practical classes*" (M = 4.08; SD = 1.32). In spite of the previous data, there is some indecision of the participants (M = 3.16; SD = 1.13) in agreeing or disagreeing with the expression "*Performing bioinformatics activities in the classroom is more time consuming*".

When asked about their initiates to explore bioinformatics tools on their own: nine out of thirteen participants reported never had independently explored bioinformatics tools; four out of thirteen reported having explored. Among the nine participants who have never independently explored bioinformatics tools in the classroom context, six of them enumerate as the main reason for this non- achievement the lack of knowledge of bioinformatics applications and their potentialities; two of them mention the complexity of the theme and one refer not to have the needed resources at school. Three out of four participants who reported that they have independently explored bioinformatics tools to implement in their classes state that they have actually explored classroom resources; one mention not to implement due to complexity of the tools explored. Eight of the thirteen participants consider that the school have the necessary conditions (computers and internet access) for the use of classroom bioinformatics tools; three mention their school have conditions and one pointed out that the needed conditions sometimes are not available.

Teachers' poor knowledge on bioinformatics was expected. When confronted with the statement': "*My initial training allows me to approach the contents to be taught through bioinformatics tools*" the participants disagree completely (M = 1.69; SD = 0.855). On the other hand, when asked to rate their agreement about the expression "*My professional training allows me to approach the contents to be taught through bioinformatics tools when they are appropriate*", the participants seem to agree that continuous training courses contribute to enrich (M = 2.77; SD = 1.24) the initial formation that participants described as weak in this area. It is worth mentioning that most of the teachers involved in this training course took their degree over a decade ago, i.e. according to a pre-bologna plan, when computer related activities were basic or inexistent. Even the teachers' that acquired recent university degrees emphasized that their training on bioinformatics at the university as undergraduates was minor and negligible for their education as teachers.

Although the participant teachers considered that the training course was well structured and organized, when questioned about the contents of the course, they pointed out : i) difficulties to interpret the bioinformatics data (3 out of 13 participants); ii) technical English / complexity of some applications / theoretical information associated (2 out of 13 participants); and iii) scientific complexity / variety of tools / personal time required (2 out of 13 participants). When prompted to suggest improvements related to the implementation of the activities of the training course, eleven out of thirteen participants did not respond. The suggestions from the two responsive participants were a deeper exploitation of the set of terms / acronyms / specific designations needed to

understand some bioinformatics exercises and more in-depth use of some bioinformatics platforms. In order to tackle some of these difficulties a website and a discussion forum was created to continuously support teachers and contribute to encourage the implementation of bioinformatics activities on teaching practices (Figure IV - 3).

Figure IV-3. Webpage: https://bioinformaticaaula.wixsite.com/bioinformatica-pt

5. Conclusion

Beyond the valuable diagnose about teacher's perceptions and knowledge on bioinformatics, this training course for in-service school teachers allowed to identify strengths and weaknesses informative to improve future editions of this training course. As main strengths, the participant teachers highlighted the curricular adequacy of the course contents; the user-friendly interfaces of the bioinformatics resources used; and their confidence to implement the bioinformatics activities in their classes. Among the main weaknesses, teachers emphasized the English-based language of bioinformatics tutorials; the multitude of commands characteristic of several bioinformatics applications; and difficulties related with data analysis and interpretation.

References

Amenkhienan, E., & Smith, E. J. (2006). A web-based genetic polymorphism learning approach for high school students and science teachers. *Biochemistry and Molecular Biology Education*, *34*(1), 30–33. https://doi.org/10.1002/bmb.2006.49403401030

European Molecular Biology Laboratory. (2017). *Bioinformatics for the classroom*. European Learning Laboratory for the Life Sciences (ELLS). http://emblog.embl.de/ells/

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. In *PLoS Computational Biology*, *7*(10), 1002243. https://doi.org/10.1371/journal.pcbi.1002243

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Marques, I., Almeida, P., Alves, R., Dias, M., Godinho, A., & Pereira-Leal, J. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Netherlands Bioinformatics Centre. (2017). *Bioinformatics@school*. http://www3.cmbi.umcn.nl/bioidklas/en/

Newman, L., Duffus, A., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

Shuster, M., Claussen, K., Locke, M., & Glazewski, K. (2016). Bioinformatics in the K-8 Classroom: Designing Innovative Activites for Teacher Implementation. *International Journal of Designs for Learning*, *7*(1), 60–70. http://www.ncbi.nlm.nih.gov/pubmed/27429860

Taylor, J., Davidson, R., & Strong, M. (2014). Drug-resistant Tuberculosis. *The American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School-New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Integrating Bioinformatics in Elementary and Secondary Education: Teachers' Perceptions

Abstract. *The new challenges in life science research highlight the need of updating science teaching practices. The aim of this study was to unveil teacher's perceptions about bioinformatics and to identify learning opportunities. In this regard, a questionnaire was applied to two different groups of secondary school level science teachers. Group A (n=11) answered the questionnaire before receiving training to implement the bioinformatics activity "Mining the Genome: using Bioinformatics Tools in the Classroom to Support Students' Discovery of Genes" in their classes. This initial survey was important to prepare teaching materials (https://bioinformaticaaula.wixsite.com/bioinformatica-pt) as well as to revise the questionnaire. An improved version of the questionnaire was applied to Group B (n=13) which included a group of teachers that attended the training course "Adequacy of Bioinformatics Tools to Elementary and Secondary Education" (CCPFC/ACC-88413/16) at Faculty of Sciences, UP. IBM SPSS was used to analyze the quantitative data, and a content analysis was carried out for qualitative data.*

All participant teachers (Groups A and B) demonstrate a high interest in bioinformatics, agreeing that it assumes a key role in biological research and highlight the need of introducing curricular framed bioinformatics activities at elementary and secondary school levels. When teachers were asked about constraints to implement bioinformatics activities in the classroom, they mainly mentioned logistics problems, namely, poor internet access and lack of computers. However, the observations recording the implementation of the bioinformatics activities in the classroom showed that these constraints were not, in fact, a limitation since the schools were equipped with computers and have internet access. Both groups agree that "The training courses available to address bioinformatics are still scarce". Adding to this, all the teachers from Group A showed interest in attending a training course in this scientific field. Regarding Group B, teachers' perceptions about their knowledge concerning bioinformatics increased during the training course. These results underline the importance of teachers' training to enhance their skills and encourage innovation in their teaching practices.

Altogether, these data concerning teachers' perceptions about bioinformatics and its curricular relevance, calls for didactics' initiatives aiming to integrate bioinformatics activities into elementary and secondary education practices.

Keywords. Bioinformatics activities, Secondary school level, Interest, Teacher Training, Teaching

1. Introduction

Nowadays, biological research is strongly benefiting from recent technological advances. In fact, the algorithms for analyzing biological data have become more sophisticated and the performance of computers are continuously improving, allowing a deeper analysis of data (National Research Council, 2005). Due to this new paradigm, the importance of bioinformatics is being highlighted among the scientific community. Bioinformatics can be defined as "*the science of how information is generated, transmitted, received, and interpreted in biological systems*" (Ramsden, 2015) or, more briefly, "the *field of study that uses computation to extract knowledge from biological data*" (Nature, 2018).

This field of research highlights the importance of updating science teaching practices to give students a glimpse of daily research routines in bioinformatics (Wefer & Sheppard, 2008). In this regard, several studies have been published underlining the benefits of including basic bioinformatics in secondary level curricula (Cummings & Temple, 2010; Machluf et al., 2017; Magana et al., 2014). Moreover, European and International initiatives have been developed to work with educational agents, policy makers and students aiming to adequate the educational resources to updated scientific practices (Faculdade de Ciências da Universidade do Porto, 2017; Marques et al., 2014; Martins et al., 2018; Netherlands Bioinformatics Centre, 2009; Wood & Gebhardt, 2013).

In fact, the role of the research institutions supporting this change is crucial. Research centers and universities have the know-how, the workforce and the facilities to provide teachers with knowledge and skills, namely, through the promotion of lectures or training courses. These initiatives contribute to rise perspectives on innovation in teaching practices by updating educators with the current biological research tools (Attwood et al., 2019; Koch & Fuellen, 2008).

Following a bottom-up approach, this study acknowledges teachers as instrumental agents of educational changes, and therefore believes that their cooperation and commitment are essential to successfully implement new activities in their classes. To mobilize teachers it is absolutely necessary to provide educational resources capable to

scaffold innovative teaching practices in order to release teachers from the cumbersome preparation of activities, namely by assessing the didactics potential of different bioinformatics tools; by designing detailed guidelines for students; and by proposing curricularly framed inquiry-based learning scenarios suitable to the implementation of bioinformatics-based exercises (Form & Lewitter, 2011; Machluf & Yarden, 2013; Wood & Gebhardt, 2013). In this regard, several curricularly framed bioinformatics activities focused on genome data mining, gene regulation, evolution, food microbiology or proteomics, have been recently proposed and validated (Arnold et al., 2017; Fernandes et al., 2014; Martins et al., 2018; Taylor et al., 2014; Wefer, 2003). In the last few years, some studies addressing the constraints to integrate basic bioinformatics-based activities in elementary and secondary school level, emphasized that teachers feedback resulting from their experience and pedagogical know-how are particularly important to properly optimize and adequate these activities in the school context (Form & Lewitter, 2011; Machluf & Yarden, 2013; Marques et al., 2014).

Regardless these contributions, teacher's perceptions about their knowledge in bioinformatics remains poorly characterized. This study focused on gathering data to characterize teacher's know-how and interest in bioinformatics, to identify current training opportunities in bioinformatics for pre-service and in-service teachers, and to make the diagnostics of weaknesses that might undermine the integration of bioinformatics in the classroom.

2. Methods

2.1. Sample and Study Context

According to the described aim of this study, two different groups of secondary school biology teachers were studied. One group had no previous training in bioinformatics but were willing to implement bioinformatics activities in their classes (Group A) and the other participated in a bioinformatics' training course (Group B) (Figure IV - 4).

Group A (n=11) included teachers who accepted to collaborate in the implementation of an activity particularly optimized for secondary level biology classes and designated as "*Mining the Genome: using Bioinformatics Tools in the Classroom to Support Students' Discovery of Genes*" (Martins et al., 2018). Group A teachers belonged to 5 schools, 3 public and 2 private, from Porto and Lisboa (Portugal), during the 2016/17 school year. After the establishment of the described collaboration and having school's directive board approval, we proceeded with the preparation of the educational resources required for the activity. To make teachers totally comfortable with the proposed materials, they were given the opportunity to make comments and suggest

improvements. Moreover, each teacher was previously trained to implement the activity, dismissing any previous doubts they had. Adding to this, teachers were given access to a dedicated webpage to support the classroom implementation of the bioinformatics activity (https://bioinformaticaaula.wixsite.com/bioinformatica-pt).

Group B included 13 in-service teachers from 12 schools (10 public and 2 private) belonging to Porto (Portugal) metropolitan area. This group of teachers participated in the first edition of the training course "*Adequacy of Bioinformatics Tools to Elementary and Secondary Education*" (2016/17), which is accredited by the competent Portuguese body (register nº CCPFC/ACC-88413/16) and which took place at Faculty of Sciences of University of Porto (Faculdade de Ciências da Universidade do Porto, 2017). The main goal of this training course is to provide teachers with the knowledge and skills to integrate bioinformatics tools and resources in their teaching practices.

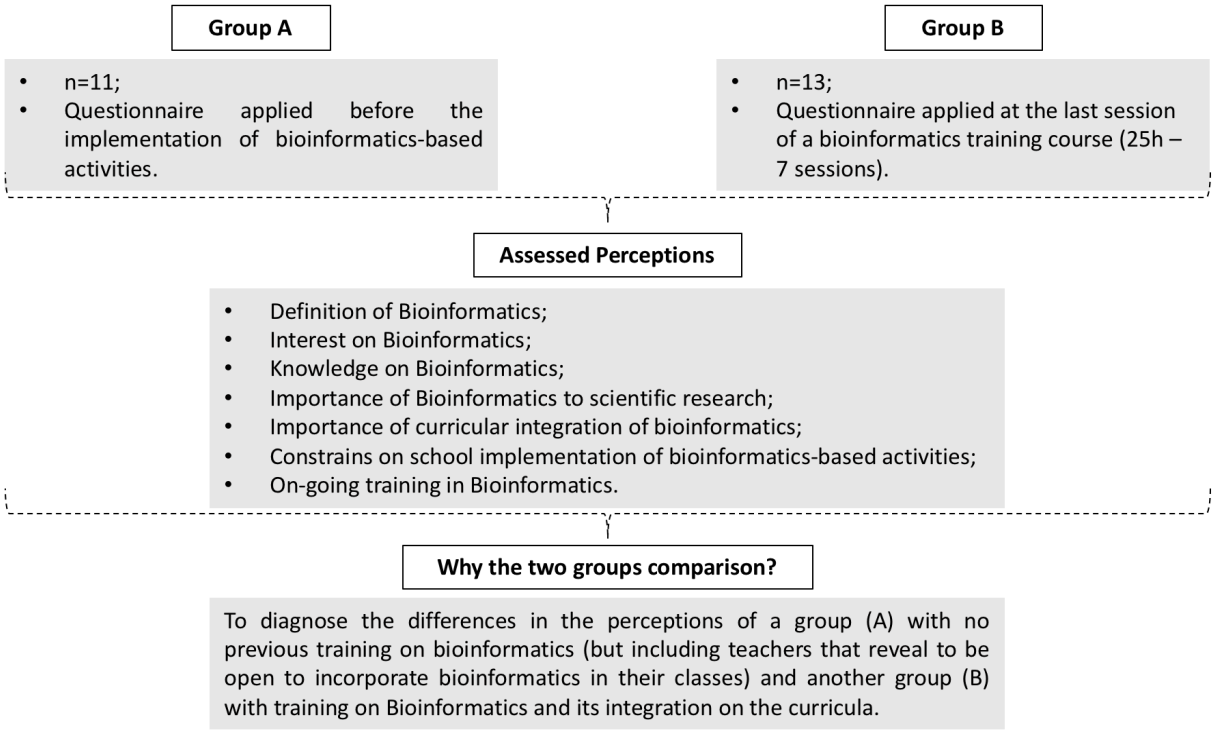


Figure IV-4. Experimental design.

2.2. Instruments

A specifically designed questionnaire was applied to both groups (A and B) (Figure IV - 5).

Part A

Name (*optional*):
**Name the institution where you obtained your Professional degree for teaching:* _____
**Year when you obtained your Professional degree for teaching:* __ __ __ __
Name the school where you are currently teaching: _____
Name the subjects you are currently teaching: _____
School level(s) you teach: _____
Contacts: Phone number _____ Email address: _____@____.____

Part B

Q1: What is *Bioinformatics* for you?

Q2: Do you think that bioinformatics-based activities are more suitable to be framed: *(a)* in the Biology curricula; *(b)* in the Information and Communications Technology (ICT) curricula; or *(c)* in both Biology and ICT curricula.

Q3: Rate your interest in bioinformatics: 1 (*Not interested at all*) – 5 (*Very interested*)

Q4: Rate your knowledge regarding bioinformatics: 1 (*Insufficient*) – 5 (*High*)

***Q4:** Rate your knowledge regarding bioinformatics: *(a)* Before the training course: 1 (*Insufficient*) – 5 (*High*); *(b)* After the training course: 1 (*Insufficient*) – 5 (*High*)

Q5: From your point of view, is bioinformatics important for the current scientific research?: 1 (*Not important at all*) – 5 (*Very important*)

Q6: In your opinion, are bioinformatics-based approaches important in the elementary school level?: 1 (*Not important at all*) – 5 (*Very important*)

Q7: In your opinion, are bioinformatics-based approaches important in the secondary school level?: 1 (*Not important at all*) – 5 (*Very important*)

Q8: Do you frequently use computers/tablets in the classroom? Yes __ No __

***Q8.1**.: If so, please indicate how frequently in a school year do you use computer resources in the classroom: 1 (*Never*) – 5 (*Very often*)

Q8.2.: If not, please indicate the main reason(s) why you did not frequently use computer/tablets in the classroom.

Q9: Have you explored bioinformatics tools by yourself in order to implement bioinformatics-based activities in your classes? Yes __ No __

Q9.1.: If so, did you implement the explored resources in the classroom? Yes __ No __

Q9.1.1.: If not, please indicate the main reasons why you do not implement the resources in the classroom.

Q10: From your point of view, please list the main constraints regarding the implementation of bioinformatics-based activities in the classroom.

Q11: Do you think that your school/institution has the needed conditions (computers and internet access) to explore bioinformatics tools in the classroom? Yes __ No __

Q12: Please rate your agreement with the following sentences - 1 (*I totally disagree*) – 5 (*I totally agree*): **(a)** My academic background allows me to approach the curricular contents using bioinformatics tools; ***(a)** My initial training at the university allows me to approach the curricular contents using bioinformatics tools; **(b)** Students are acquainted with bioinformatics and its resources; ***(b)** My professional training allows me to approach the curricular contents using bioinformatics tools; **(c)** Preparing a class using bioinformatics tools requires more time and resources than preparing other practical classes; **(d)** Performing a class using bioinformatics tools is more time consuming than preparing other practical classes; **(e)** Training courses for teachers in bioinformatics are still scarce.

Q13: Would you be interested in attending more training courses/workshops promoted by research groups which use bioinformatics tools in their lab routines? Yes __ No __

***Q13:** Please list the main difficulties that you found while performing the activities proposed in the training course.

***Q14:** Please make suggestions for improvements that you consider important concerning the activities of the training course you are attending.

****Q15 = Q13**

****Q16:** Do you accept the inclusion of your email in the mailing list of MDE/CIBIO-InBIO and FCUP to receive information regarding future training courses?

Figure IV-5. Questionnaire used in the study. Items highlighted green (*) were rephrased taking into account Group A teacher's feedback to improve the comprehension of these items and were included in the questionnaire given to Group B teachers. Items highlighted blue (**) were added to the questionnaire given to Group B teachers.

The questionnaire includes three parts. Part A was aimed to collect data regarding socio demographic characterization of the sample. Part B covered the assessment topics regarding teachers' university training and academic background, their attitudes towards bioinformatics and their perceptions about the availability of bioinformatics resources. Part C included three additional items to assess teacher's opinions about the questionnaire objective, the comprehension of the items, and an open space to make suggestions for improvement (these items are not displayed in Figure IV - 5).

2.3. Data Collection

The initial version of the questionnaire was applied to Group A before this group got training and support to implement the bioinformatics activity in their classes. Teachers answered the questionnaire during a get-together session involving all the collaborating teachers *per* school. This survey was important to improve the questionnaire as a data collection instrument. The improved version of the questionnaire, which included new (**Q13, **Q14 and **Q16) and rephrased items (Q4/*Q4, *Q8.1. Q12 – (a), *(a), (b), *(b)), was applied to Group B during the last session of the training course. It is important to add that the feedback from Group A teachers gave an important contribution to prepare teaching materials, which were made available in the "*Teaching Resources*" section of the webpage *Bioinformática na Sala de Aula* (https://bioinformaticaaula.wixsite.com/bioinformatica-pt).

2.4. Data Analysis

Data analysis included the use of quantitative and qualitative techniques using a mixed-method approach to increase the consistency of the analysis (Punch, 2009). IBM SPPS version 24 was used to analyze the quantitative data. Descriptive statistics included the non-parametric tests Mann-Whitney Test and Wilcoxon Signed Rank Test (Pallant, 2007). Qualitative data was analyzed according to content analysis procedures. In the open-access answers (Q1, Q8.2., Q9.1.1., Q10, **Q13, and **Q14) a content cloud analysis was carried out to have a quick view of the most frequent notions mentioned. This technique was performed using *Wordle* (available at: http://www.wordle.net/) as a research adjunct tool of research according to previous studies (Cidell, 2010; McNaught & Lam, 2010).

3. Results and Discussion

This research was based on an experimental design, detailed in Figure IV - 4, conceived to optimize an instrument to assess teachers' perceptions about bioinformatics. Most importantly, this study gathers data regarding teachers' perceptions

and interest about bioinformatics, contributing to orient new teaching practices and to identify learning opportunities for secondary education in this field.

3.1. An Instrument to Assess Teachers' Perceptions about Bioinformatics

The questionnaire given to teachers from Group A allowed to collect data concerning this instrument in order to make an improved version with new (**Q13, **Q14 and **Q16) and rephrased items (Q4/*Q4, *Q8.1. Q12 – (a), *(a), (b), *(b)), focused on questionnaire comprehension and to further potentiate its utility for future assessments. Regardless the modifications based on teachers' critical appraisal of the questionnaire, the participants of Group A considered the instrument objective (M=4.45; SE=0.21) and easy to understand (M=4.82; SE=0.12) in a five-point Likert-type scale, therefore validating its utility as an assessment tool of teachers' perception about bioinformatics.

Four participants mentioned the need to change the expression *academic background* in the sentence "*My academic background is adjusted to the requirements to teach using bioinformatics tools, when adequate*" (see Figure IV - 5 – Q12 – *(a)*). The term was replaced by *initial training at the university* (see Figure IV - 5 – Q12 – **(a)*). Furthermore, it was acknowledged the comments made by 4 teachers who highlighted the importance to distinguish their university training (i.e. pre-service), which for some was over two decades ago, from the attendance of training courses and workshops they got more recently, i.e. as in-service teachers. In this regard a new item was added: "*My professional training is adjusted* to the requirements to teach using bioinformatics tools, when adequate" (see Figure IV - 5 – Q12 – **(b)*).

Three participants suggested adding a Likert-type scale to the question: "*Please indicate how regularly in a school year do you use technology in your class activities*", since it was easy for them recall these teaching interventions, and simultaneously help to quantitatively measure this item. The Likert-type scale added was from 1 (Never) to 5 (Very often) (see Figure IV - 5 – Q8 – **Q8.1.*).

Two participants referred that open-ended questions (Q1, Q8.2., Q9.1.1., and Q10) should be reduced or, alternatively, replaced by multiple choice questions. Considering that the number of "*no answers*" was low ("No answer" frequency: 3/11) and that all the participants of Group A answered the questionnaire on time and easily, we decided to keep the number of open-ended questions of the questionnaire. In addition, most of these open-ended questions revealed to be particularly useful for a qualitative assessment through content analysis of teachers' knowledge (Q1); their opinion about

the curricular framing of bioinformatics activities (Q2); and the constraints to implement bioinformatics exercises in classroom (Q10).

Interestingly, one participant from Group A, who was particularly concerned with paper waste that these questionnaires may imply, suggested the adoption of a digital version of the questionnaire. Although proposed as an eco-friendly solution, the fact is that a digital version might contribute to rapidly increase the sampling universe, thus permitting more robust assessments. This was a particularly pertinent suggestion that will be taken into account in further studies.

Four participants referred not to have suggestions for improvements.

After analyzing the suggestions and the answers of the Group A version of the questionnaire, it was also decided to make the following extra changes to better characterize the participant teachers:

Sample characterization data (Figure IV - 5 – Part A – **highlighted information*): it was added a blank space to indicate the year when teachers obtained their professional degree for teaching. This information is important to better characterize their teaching experience and simultaneously to gather data about the impact their academic training had on a relatively new field that is bioinformatics. Some questions particularly dedicated to teachers that attended the training course (i.e. Group B teachers) were added to the questionnaire (***Q13, **Q14, **Q16*). These questions aimed to improve the training formation in order to better meet teachers' needs regarding the inclusion of bioinformatics activities in their teaching.

3.2. Findings on Teachers' Perceptions

Bioinformatics Definition and Importance

Regarding the definition of bioinformatics (Q1), the teachers from both Groups A and B revealed to have a correct notion when compared with definitions from experts of this scientific field (Nature, 2018; Ramsden, 2015). A content cloud analysis, limited to 20 words, taking together the answers of the two groups (A and B), reveals that, in general, all participants relate bioinformatics with the use of computers, which they considered to be crucial to analyze rapidly and effectively biological data sets (Figure IV - 6). This result suggests that, even without bioinformatics training, the teachers are well acquainted with bioinformatics as an important tool to obtain meaningful interpretation of biological results.

Figure IV-6. Content cloud of teachers' definitions of bioinformatics (analysis limited to 20 words, participants of Group A and Group B were considered together, n=24).

This result is likely related with teachers' interest in this field (Q3). In fact, it can be concluded that teachers of both groups revealed to have a high interest in bioinformatics and were motivated to explore this area (Figure IV - 7) (U = 60.00, z = -0.76, p = 0.45, r = 4.9). Moreover, teachers agree on the importance of bioinformatics for current scientific research (Q5) (U = 66.00, z = -0.92, p = 0.36, r = - 0.19). These findings reinforce previous studies which mention the motivation of teachers to learn more about bioinformatics in order to engage their students in up-to date scientific areas (Kovarik et al., 2013; Wood & Gebhardt, 2013) (Figure IV - 7).

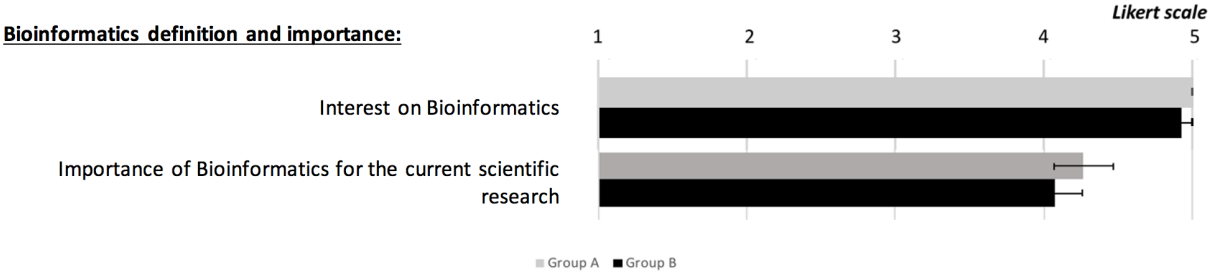


Figure IV-7. Assessment of teachers' interest in bioinformatics and its importance for scientific research. Bars represent mean with standard error.

Teachers' Perception about their Knowledge of Bioinformatics

Group A teachers tended to disagree that their academic background (Q12 - a) allows to approach the curricular contents using bioinformatics tools. In Group B, although the *initial training* and *professional training* were analyzed separately (Q12 - *a; Q12 - *b), the results were not significantly different from the ones obtained with Group A (Figure IV - 8).

Participants of Group A and Group B admitted being insecure about their knowledge in bioinformatics before the training, i.e. for Group A before implementing the bioinformatics activity "Mining the Genome", and for Group B before the training course (Q4; *Q4 – a) (Figure IV - 8). This lack of confidence has been previously reported and highlights the need to train teachers in this field (Dalpech, 2006; Kovarik et al., 2013;

Wood & Gebhardt, 2013). Furthermore, the bioinformatics training course reported by Martins, Lencastre, & Tavares (2017) revealed to be an important contribution to nurture the enthusiasm of participant teachers about bioinformatics, who end up feeling more confident to implement bioinformatics-based activities in their classes. These results are aligned with a recent study which reports that after a first foray into bioinformatics activities, teachers tend to feel more motivated, confident and able to redesign the didactic-pedagogical goals (Machluf et al., 2017). Not surprisingly, Group B perception of their knowledge on bioinformatics was significantly different after the course training sessions (*Q4 – b) (z = -3.33, p < 0.01, r = - 0.64), further confirming the positive impact on teachers' knowledge but also on their confidence to tackle these matters during their teaching practice.

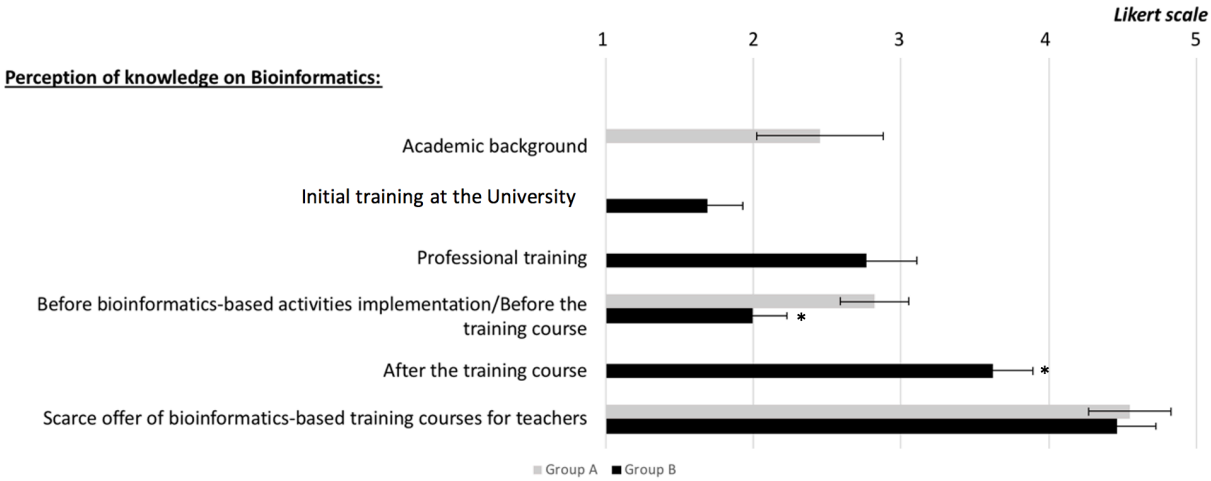


Figure IV-8. Assessment of teachers' perceptions regarding their knowledge on bioinformatics. Bars and error bars represent mean +/- standard error. *Statistically significant different (p<0.01): Group B Before the training course x Group B After the training course.

Knowing that all participants revealed to use computers in their classes (Q8), the two groups were asked about their autonomy to explore and apply bioinformatics tools in their classes (Q9). Interesting, 8 participants from Group A revealed to have autonomously used bioinformatics tools against only 4 teachers in Group B. This result suggests a different profile between the two groups. Teachers of Group A, who decided to collaborate with us to implement in their classrooms the activity "Mining the Genome" described by Martins, Fonseca & Tavares (2018), without previous training, revealed a spontaneous initiative to search and explore bioinformatics platforms by themselves. On the contrary, Group B teachers, who apply for a course training in bioinformatics, includes mostly teachers who do not felt this previous confidence and decided to have insights in this field.

Teachers' search for bioinformatics training contrasts with the information given by all the participants (Group A and Group B) about the scarce offer of bioinformatics-based training courses for secondary school teachers (Q12 – e) (Figure IV - 8). The emphasis

on the need for more training courses for in-service teachers in bioinformatics has been previously reported, underlying its contribution to enrich the initial academic training described by participants as weak (Martins et al., 2017; Ranganathan, 2005; Schneider et al., 2010). When considering both groups together, only 1 participant showed no interest in participating in training courses promoted by research and academic experts who routinely use bioinformatics tools (Q13).

Adequacy of Bioinformatics-based Activities to the Educational Context

The majority of participants (8 of Group A and 10 of Group B) consider that bioinformatics fits the biology curricula of secondary level education (Q2). Six teachers consider that bioinformatics can be helpful for both Biology and Information and Communications Technology (ICT) classes of Portuguese secondary level curricula. In fact, bioinformatics is an interdisciplinary area of knowledge integrating biology, mathematics, statistics, chemistry and computer sciences, which makes this scientific discipline particularly suitable to different classes of secondary education. The bioinformatics activity "Mining the Genome" (for Group A teachers) and the bioinformatics exercises of the training course (for Group B teachers) were particularly designed for Biology teachers based on a question-driven approach to address issues such as gene regulation and evolution (Martins et al., 2018), or real-world problems concerning food preservation and safety (Fernandes et al., 2014).

Interestingly, although the bioinformatics activities supporting this study were particularly framed for the secondary level curricula of Biology, the participants of both groups (Group A and Group B) considered that bioinformatics is important for secondary school level (Q7) (U = 58.00, z = -0.91, p =0.36, r = - 0.19), as well as for the elementary school level (Q6) (U = 41.5, z = -1.37, p = 0.17, r = - 0.29) (Figure IV - 9). It is important to emphasize that Group B teachers had a clearer vision on how to apply bioinformatics tools in the classroom, suggesting a careful reflection regarding the learning utility of the bioinformatics exercises proposed in the training course and its integration with the curricular contents. On the contrary, this effort was previously carried out by us by making ready to use materials, such as guidelines to students and power point presentations curricularly framed when the activity "Mining the Genome" was proposed to Group A teachers. The leaning importance given to bioinformatics by both groups reinforces the interest of the theme, even taking into account that presently this scientific field is generally absent from the secondary and elementary school curricula worldwide (Wefer & Sheppard, 2008; Wood & Gebhardt, 2013).

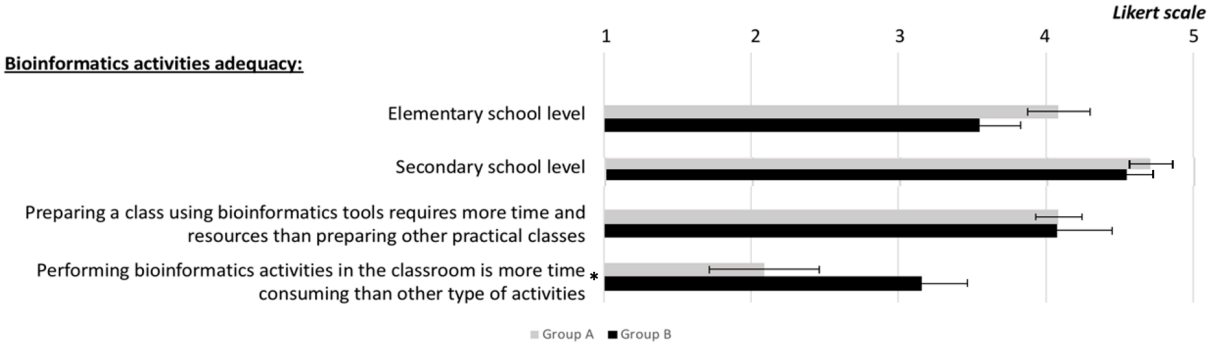


Figure IV-9. Assessment of teachers' perceptions about bioinformatics adequacy to the educational context. Bars represent mean with standard error. *Statistically significant different ($p<0.01$): Group A Performing bioinformatics activities in the classroom is more time consuming than other type of activities x Group B Performing bioinformatics activities in the classroom is more time consuming than other type of activities.

During this study it was crucial to have an insight on teachers' feedback about the available school resources to implement bioinformatics-based activities in the classroom. The majority of the teachers from Group A (8 out of 11) agreed that their school have the necessary conditions to implement the activity "Mining the Genome" (Q11). In Group B, which was mainly composed with teachers who did not had the opportunity to apply the bioinformatics activities learned during the training course in their classes, are more skeptical about the school resources required to implement bioinformatics exercises. In fact, only 3 out of 13 teachers from Group B considered that schools have the facilities needed for the successfully implementation of bioinformatics activities. The main constraints mentioned by teachers of both groups were coincident (Q9.1.1.) (Figure IV - 10).

Group A *Group B*

Figure IV-10. Content cloud of mentioned constraints to apply bioinformatics in the classroom. (limited to 10 words).

The constraints most frequently mentioned were logistics issues, namely, poor internet access, lack of computers and of time to prepare the classes, which were somewhat expected limitations according to the conclusion of other studies on bioinformatics-based initiatives (Marques et al., 2014; Wood & Gebhardt, 2013). However, the observation record carried out by us when visiting the schools from Group A, showed that these constraints were not impairing the implementation of the bioinformatics activity. Some European studies have been carried out in order to assess the integration of ICT in education and to identify improvements and constraints in the

implementation of ICT based approaches (Korte & Hüsing, 2007; Snyder & Dillow, 2013; Wastiau et al., 2013). Pointing to a number of policy actions, it is also reported that nowadays schools are generally well-equipped in what concerns to computers availability and internet connection. This information suggests a possibility to implement bioinformatics-based activities in the classroom.

Regarding time constraints to prepare and implement bioinformatics activities to teach science topics, and despite teachers' acknowledgment that bioinformatics might be useful in their teaching practices, teachers of both groups agree that preparing a class using bioinformatics tools requires more time and resources than preparing other practical classes (Q12 – c) (U = 57.50, z = -0.88, p =0.38, r = - 0.18) (Figure IV - 9). This result is in line with other studies concerning the implementation of innovative practical approaches in the classroom, namely involving technology to approach scientific issues, as for example biotechnology (Fonseca et al., 2012; Martins et al., 2017). The strict teaching schedule, mainly caused by the outsized curricular contents, is pointed out by teachers as a major reason to not diversify the didact approaches in the classroom (Bryce & Gray, 2004; Fonseca et al., 2012). In contrast, when asked if carrying out the bioinformatics activities in the classroom is more time consuming than other type of activities (Q12 – d) the teachers of both groups have a significantly different profile, with Group A teachers being more optimist than Group B (U = 27.5, z = -2.62, p <0.01, r = - 0.53) (Figure IV - 9). Again, the previous preparation of the teaching materials given to Group A teachers and the factual assistance to implement the activity "Mining the Genome", may have strongly influence their positive attitude regarding the time required to complete the activity in the classroom. On the other hand, although Group B teachers got the training, the basic knowledge and competences to use user-friendly and open access bioinformatics tools in their educational practices (Faculdade de Ciências da Universidade do Porto, 2017; Martins et al., 2017), the fact is that these teachers still have to prepare the teaching materials and are unsure about students' performance when facing a new work challenge. The bioinformatics training course attended by Group B teachers was expected to divulge user-friendly bioinformatics resources suitable to be rapidly implemented in the classroom, as suggested by other studies which reported that the majority of the teachers implement hands-on bioinformatics-based activities when have access to them (Shuster et al., 2016; Wood & Gebhardt, 2013). Moreover, when teachers are supported by appropriated training, they feel skilled to create their own bioinformatics research-driven exercises, claiming the positive impact of these approaches in their students' learning (Shuster et al., 2016; Wood & Gebhardt, 2013). Indeed, Group B teachers had no difficulties when challenged to prepare a bioinformatics

activity for their classes. Regardless all these training efforts, Group B teachers' perception regarding the time required to implement some of the learned bioinformatics exercises was not as anticipated, suggesting that providing adequate teaching materials might be essential to raise teachers' receptivity to effectively apply bioinformatics activities in the classroom.

4. Conclusion

This study contributed to characterize teachers' perceptions about bioinformatics and to stress its relevance as a didactic tool. The results highlighted are important indicators to foster educational updates aiming to integrate bioinformatics in the elementary and secondary school curricula. Extending this study to a larger universe of biology teachers will be important to consolidate the indicators. Most importantly, in order to promote bioinformatics potential as a teaching tool, more training courses and workshops on how to prepare the teaching materials will be needed. It is expected, that these and future contributions, might facilitate a bottom-up approach that ultimately might lead to gradually integrate basic bioinformatics contents in science curricular orientations for elementary and secondary level.

Acknowledgments

The authors are grateful to all participants of this study (teachers and schools) and to Leonor Martins for the fruitful comments made on the manuscript. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Arnold, M. L., Holman, D., & Zweifel, S. G. (2017). Using Molecular Biology and Bioinformatics to Investigate the Prevalence of Mislabeled Fish Samples. *The American Biology Teacher*, *79*(9), 763–768. https://doi.org/10.1525/abt.2017.79.9.763

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., & Schneider, M. V. (2019). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*, *20*(2), 398–404. https://doi.org/10.1093/bib/bbx100

Bryce, T., & Gray, D. (2004). Tough acts to follow: The challenges to science teachers presented by biotechnological progress. *International Journal of Science Education*, *26*(6), 717–733. https://doi.org/10.1080/0950069032000138833

Cidell, J. (2010). Content clouds as exploratory qualitative data analysis. *Area*, *42*, 514–523. https://doi.org/10.2307/40890909

Cummings, M. P., & Temple, G. G. (2010). Broader incorporation of bioinformatics in education: Opportunities and challenges. *Briefings in Bioinformatics*, *11*(6), 537–543. https://doi.org/10.1093/bib/bbq058

Dalpech, R. (2006). Bioinformatics and school biology. *Journal of Biological Education*, *40*(4), 147–148. https://doi.org/10.1080/00219266.2006.9656035

Faculdade de Ciências da Universidade do Porto. (2017). *Adequação de ferramentas bioinformáticas ao 3ºciclo do Ensino Básico e ao Ensino Secundário*. Cursos de Formação Contínua. https://sigarra.up.pt/fcup/pt/cur_geral.cur_view?pv_ano_lectivo=2017&pv_origem=CUR&pv_tipo_cur_sigla=CFC&pv_curso_id=13321

Fernandes, E., Dias, C., Fonseca, M. J., & Tavares, F. (2014). Understanding Growth and Thermal Inactivation of Foodborne Bacteria Using the Pathogen Modelling Program (PMP). In M. Costa, P. Pombo, & J. Dorrio (Eds.), *Hands-on Science: Science Education with and for Society* (pp. 207–210). Hands-on Science Network.

Fonseca, M. J., Costa, P., Lencastre, L., & Tavares, F. (2012). Disclosing biology teachers' beliefs about biotechnology and biotechnology education. *Teaching and Teacher Education*, *28*(3), 368–381. https://doi.org/10.1016/j.tate.2011.11.007

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. *PLoS Computational Biology*, *7*(10), e1002243. https://doi.org/10.1371/journal.pcbi.1002243

Koch, I., & Fuellen, G. (2008). A review of bioinformatics education in Germany. *Briefings in Bioinformatics*, *9*(3), 232–242. https://doi.org/10.1093/bib/bbn006

Korte, W. B., & Hüsing, T. (2007). Benchmarking Access and Use of ICT in European Schools 2006: Results from Head Teacher and A Classroom Teacher Surveys in 27 European Countries. *eLearning Papers, 2*(1).

Kovarik, D. N., Patterson, D. G., Cohen, C., Sanders, E. A., Peterson, K. A., Porter, S. G., & Chowning, J. T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Machluf, Yossy, Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-

school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Magana, A. J., Taleyarkhan, M., Alvarado, D. R., Kane, M., Springer, J., & Clase, K. (2014). A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE—Life Sciences Education*, *13*(4), 607–623. https://doi.org/10.1187/cbe.13-10-0193

Marques, I., Almeida, P., Alves, R., Dias, M. J., Godinho, A., & Pereira-Leal, J. B. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r1_id=

McNaught, C., & Lam, P. (2010). Using Wordle as a Supplementary Research Tool. *The Qualitative Report*, *15*(3), 630-643. https://nsuworks.nova.edu/tqr/vol15/iss3/8

National Research Council. (2005). *Catalyzing Inquiry at the Interface of Computing and Biology* (J. Wooley & H. Lin (eds.)). The National Academies Press.

Nature. (2018). *Bioinformatics*. Nature.Com. https://www.nature.com/subjects/bioinformatics

Netherlands Bioinformatics Centre. (2009). *Bioinformatics as a didactic tool in high school*. NBIC Teacher Training. https://www.nbic.nl/education/high-school-programmes/bioinformaticsschool/teacher-training/index.html

Pallant, J. (2007). *SPSS - Survival Guide to Data Analysis using SPSS for Windows*. Open University Press/McGraw-Hill.

Punch, K. F. (2009). *Introduction to Research Methods in Education.* SAGE Publications.

Ramsden, J. (2015). *Bioinformatics: An Introduction* (Vol. 21). Springer. https://doi.org/10.1007/978-1-4471-6702-0_1

Ranganathan, S. (2005). Bioinformatics Education—Perspectives and Challenges. *PLoS Computational Biology*, *1*(6), e0010052. https://doi.org/10.1371/journal.pcbi.0010052

Schneider, M. V., Watson, J., Attwood, T., Rother, K., Budd, A., McDowall, J., Via, A., Fernandes, P., Nyronen, T., Blicher, T., Jones, P., Blatter, M.-C., De Las Rivas, J., Judge, D. P., van der Gool, W., & Brooksbank, C. (2010). Bioinformatics training: a review of challenges, actions and support requirements. *Briefings in Bioinformatics*, *11*(6), 544–551. https://doi.org/10.1093/bib/bbq021

Shuster, M., Claussen, K., Locke, M., & Glazewski, K. (2016). Bioinformatics in the K-8 Classroom: Designing Innovative Activites for Teacher Implementation. *International Journal of Designs for Learning*, *7*(1), 60–70. http://www.ncbi.nlm.nih.gov/pubmed/27429860

Snyder, T., & Dillow, S. (2013). *Digest of Education Statistics 2012 (NCES 2014-015)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Taylor, J. M., Davidson, R. M., & Strong, M. (2014). Drug-resistant Tuberculosis. *The American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Wastiau, P., Blamire, R., Kearney, C., Quittre, V., Van de Gaer, E., & Monseur, C. (2013). The Use of ICT in Education: A survey of schools in Europe. *European Journal of Education*, *48*(1), 11–27. https://doi.org/10.1111/ejed.12020

Wefer, S. H. (2003). Name That Gene: An Authentic Classroom Activity Incorporating Bioinformatics. *The American Biology Teacher*, *65*(8), 610–613. https://doi.org/10.2307/4451571

Wefer, S. H., & Sheppard, K. (2008). Bioinformatics in high school biology curricula: A study of state science standards. *CBE Life Sciences Education*, *7*(1), 155–162. https://doi.org/10.1187/cbe.07-05-0026

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School-New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers

Abstract. *Having in mind the technological world we live in, scaffolding teachers to integrate computer-based resources in their classes is an important endeavour when it comes to teacher professional development. In this context, a workshop on bioinformatics to address gene regulation, molecular biology and genomics in high school gathered a group of 18 high school science teachers. Following the workshop, the participants answered a questionnaire aiming to diagnose teachers' perceptions about bioinformatics; to identify the main constraints that are preventing teachers from successfully implementing bioinformatics in their classes; and to evaluate the potential to integrate basic bioinformatics exercises in their educational practices. Data were subjected to descriptive statistics and to content analysis. The results showed that the teachers attending the workshop were highly motivated and interested in learning more about bioinformatics and about strategies to integrate bioinformatics in their classes. Despite teachers highlighted the adequacy of bioinformatics to the educational context, most of them mentioned that their academic background was not sufficient to confidently implement bioinformatics-based exercises in their classes. Teachers claimed for more training courses in this area and approximately half of the participants admitted that schools are equipped with the necessary resources to integrate bioinformatics. Overall, this study emphasizes the importance to foster more initiatives to integrate bioinformatics in secondary education curriculum and highlights the need to increase the offer of teachers' training on bioinformatics.*

Keywords. Computer Based Learning, Teacher Professional Development, Technology in Education and Training

1. Introduction

Despite technology is now acknowledged as a helpful tool for teaching and learning science, many science teachers got their academic qualifications for teaching at a time in which technology, namely computers, were rarely used or inaccessible for most of the people. It is therefore with no surprise that in-service teachers feel uncomfortable or not motivated to use computers in their classrooms, being urgent to scaffold teachers to

develop their Technological Pedagogical Content Knowledge (TPACK) (Koehler & Mishra, 2000; Mumtaz, 2000).

Having this need in mind, professional development programs for science teachers should focus on helping them to integrate the use of the technology in the teaching practices. An opportunity to integrate computers and digital resources as a didactic tool based on real research procedures is to approach genomics and molecular biology based-on bioinformatics tools (Chiovitti et al., 2019; Gelbart & Yarden, 2006; Martins, Fonseca, et al., 2018; Nunes et al., 2015). Despite genomics being one of the most important revolutions of late 20th century and at the beginning of this century, the lack of literacy on genomics and molecular biology in the general population is reported in several studies (Eklund et al., 2007; Kirkpatrick et al., 2002; Kolarova, 2011; Lock, 1996; Tarver, 2010). While an effort is being made to integrate bioinformatics in high school science curricula in different countries, it is unquestionable that more teacher's training opportunities in bioinformatics is needed (Attwood et al., 2017; Kovarik et al., 2013; Machluf et al., 2017; Marques et al., 2014).

In this context, the workshop "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the Classroom*" was aimed to instruct science teachers about the suitability of bioinformatics activities to their teaching practices. Teachers were challenged to explore bioinformatics-based exercises particularly chosen to teach basic molecular biology concepts, to address gene regulation and to discover the usefulness of comparative genomics (Martins, Fonseca, et al., 2018). This training workshop on bioinformatics allowed to boost teachers' TPACK, since teachers developed their technological knowledge by learning how to use bioinformatics tools and, concomitantly, they enlarged their pedagogical and content knowledge, by discussing new strategies to teach curricular contents and additionally introduce key concepts in genomics such as Open Reading Frame (ORF) or Basic Local Alignment Tool (BLAST) (Martins & Tavares, 2018). The workshop was promoted in the context of an international meeting for teachers (Casa das Ciências, 2018) that occurred in Portugal in 2018.

2. Objectives

The main objectives of this study were to diagnose teachers' perceptions about bioinformatics; to infer the potential of its integration in educational practices; and to identify the main constraints that are preventing teachers to successfully implement bioinformatics in their classes.

3. Research Questions

This study was driven by two main research questions: "*Which are the teachers' perceptions about bioinformatics and its integration in science teaching practices*?" and "*Which are the main constraints that are preventing teachers from integrating bioinformatics in their teaching practices*?"

4. Methods

4.1. Participants

The sample included 18 science teachers (14 female and 4 male) from 14 different schools that voluntarily enrolled at the workshop. Seven of the 18 teachers hold a master's degree. Participants have an average of 26.61 ± 7.48 years of teaching experience. Between 2016 and 2018, 4 teachers taught at elementary school level (students between 12-15-year-old), 6 taught at secondary school level (students between 16-18 years old) and 8 taught both elementary and secondary school levels.

4.2. Materials

Framed within the curriculum for biology in secondary education (National Research Council, 2013), the four hours workshop "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the Classroom*" drives teachers to explore the potential of bioinformatics as a didactic tool to approach contents, such as the organization and regulation of genetic material, and to carry out evolutionary inferences. From an *in silico* analysis of a DNA sequence it is proposed to identify genes and to determine the putative functions of their products. Additionally, using comparative genomics platforms such as *MaGe – Magnifying Genomes* (MicroScope), the participants were challenged to evaluate the presence of certain genes in different taxonomic groups aiming to infer evolutionary relations. This activity contributes to a holistic approach to genomics, genes and proteins, as well as to propose evolutionary hypotheses (Martins, Fonseca, et al., 2018). Each teacher was asked to bring their personal computers to carry out the exercises.

To diagnose teachers' perceptions about bioinformatics and its integration in science teaching practices as well as to identify the main constraints that are preventing teachers from integrating bioinformatics in their classes, a specifically designed questionnaire, adapted from the survey previously described by Martins, Lencastre and Tavares (2018), was implemented. The questionnaire, including 35 questions in various formats, namely dichotomous, Likert-type (ranged from 1 to 5) and open-ended questions, and is divided

into three parts: *Part A*: socio demographic data; *Part B*: assessment of teachers' training and academic background on bioinformatics, and appraisal of teachers' attitudes towards bioinformatics integration and of their perceptions regarding workshop attendance; *Part C*: teacher's opinions about the questionnaire: objectivity, comprehension of the items, and suggestions.

In Part B, questions were designed according to the objectives defined for this study. Questions 1 and 5 (Q1; Q5) were aimed to diagnose teachers' definition of bioinformatics as well as to characterize the importance of bioinformatics to the current research. The assessment of teachers' interest and perceived knowledge about bioinformatics were addressed through questions Q3, Q12.1., Q12.2. Another dimension evaluated was teachers' perspectives on the importance of integrating bioinformatics in their teaching practices as well as to identify the main obstacles to this integration. In this regard, questions Q2, Q6, Q7, Q9, Q9.1., Q9.1.1., Q10, Q11, Q12.3., Q12.4. were included in the questionnaire. Question Q8 was intended to characterize the use of technology in the classroom by teachers. Lastly, and having in mind the main findings of previous studies highlighting the importance of promoting teachers training actions in the area of bioinformatics (Machluf et al., 2017; Machluf & Yarden, 2013; Martins et al., 2017; Martins, Lencastre, et al., 2018), questions Q4.1., Q4.2., Q12.5., Q13, Q14, Q15, Q16 and Q17 were introduced to characterize the impact of the workshop on teachers perceived knowledge about bioinformatics as well as to evaluate the workshop, identifying the potential of the action but also having feedback on possible improvements.

Teachers rated the questionnaire as an objective instrument (4.76±0.44) and easy to understand (4.76±0.44). One participant added the following suggestion for improvement: "*In question 12.4., you should specify if the classes are lectures or if they are practical classes of biology including wet lab and experiments*". This suggestion will be taken into consideration in future questionnaires.

4.3. Data Collection and Analyses

The questionnaire was implemented after the attendance of the workshop, i.e. when participants finished the proposed exercises. The aims of the research as well as the objectives of the questionnaire were explained to teachers who voluntarily agreed to answer the survey. Descriptive statistical analysis was performed for quantitative data (Punch, 2009). For qualitative data, a thematic content analysis of the participants' responses to open-ended questions was carried out (Roberts, 2015; Schuster & Weber, 2006).

5. Results and Discussion

5.1. Teachers' Background on Bioinformatics

The majority of the teachers (94.44%) correctly defined bioinformatics (Q1), and one teacher mentioned that bioinformatics is "*a didactic tool for science classes*". Among all listed notions (49), teachers mentioned frequently the following: informatics (12.62%), biology (8.74%), data (6.8%), tools (6.8%) and applications (4.85%). This analysis revealed that teachers recognized the scope of bioinformatics as the scientific field which develops or uses tools and applications of informatics to understand the biological data, which fits well with a general definition of bioinformatics (Luscombe et al., 2001; Sadek, 2004).

Teachers revealed to be interested in bioinformatics (Q3) and recognized its importance for scientific advances (Q5) (Figure IV - 11). However, teachers considered that their academic background is not sufficient to feel prepared to teach using bioinformatics tools (Q12.1.) and highlighted the added-value of professional training (in-service) (Q12.2.) to implement bioinformatics activities in their classes (Figure IV - 11). These results are in line with the literature (Machluf, et al., 2012; Machluf & Yarden, 2013; Martins, Lencastre, et al., 2018; Wood & Gebhardt, 2013) and reinforce the importance of promoting initiatives of professional development oriented for bioinformatics training.

5.2. Attitudes towards Bioinformatics Integration

The importance of integrating bioinformatics in elementary (Q6) and secondary education (Q7) was highlighted by the participants (Figure IV - 11), as well as the potential to use bioinformatics both at Biology classes and Information and Communications Technology (ICT) classes (Q2), mentioned by 77.78% of the teachers. It is acknowledged that bioinformatics-based activities foster students' hybrid abilities in computation and biology, nurturing a wide range of skills such as using bioinformatics to find, retrieve and organize data by identifying an appropriate data repository, to understand evolutionary related processes or to develop their critical thinking namely in which concerns open data access (Foster & Sharp, 2007; Mariano et al., 2019; Oliver et al., 2012; Sayres et al., 2018).

The majority of teachers (94.44%) assumed to use computers to explore online resources in their classes (Q8), but only 5 out of 18 (27.78%) revealed to have autonomously explored bioinformatics resources in order to implement them in their classes (Q9). Among the teachers who had previously explored bioinformatics

resources, 3 out of 5 (60%) actually implemented the activities in their classes (Q9.1.). The two teachers who did not (40%), listed as the main reasons "*Lack of computers available*" and "*The need to better understand the explored bioinformatics tools".* However, their intention is *"to apply the activities in the classroom soon*" (Q9.1.1.).

Around half of the participants (55.56%) considered that the school/institutions where they were teaching have the necessary conditions to integrate bioinformatics-based activities in their classes (Q11), which is in agreement with recent reports indicating that most schools in Europe are equipped with technological devices (European Comission, 2013). In contrast with this result, the main constraints identified by teachers to carry out the implementation of bioinformatics activities in the classroom (Q10) were: computers (7.87%) and internet (6.74%). Only 2 out of 18 teachers (11.11%) revealed positives attitudes regarding the possible constraints that can arise when implementing bioinformatics in their classroom mentioning that "*cannot identify any constraints. The school has the necessary resources (…) and students are motivated*" and "*the management of the time is possible especially in the 11th grade*". 66.67% of the participants indicated that the main constraints to implement bioinformatics-based activities in their classrooms were related with logistics aspects (resources and time) answering that "*computers are lacking in schools*"; "*the number of students per class is too high*"; "*the internet connection is weak*"; and "*bioinformatics resources are not in Portuguese*". Teachers mentioned that schools are well prepared to carry out bioinformatics exercises, however logistics aspects were mentioned as the main hinder to not apply these tools in the classrooms. This result stresses the importance to better characterize schools' reality in which concerns technology use.

16.67% of the participants identified both logistic constraints and lack of teacher's confidence as the main difficulties to implement the proposed approached in the classroom. Teachers emphasized their lack of confidence to approach some curricular topics using bioinformatics resources and highlighted their need to acquire specific training, which is corroborated by other studies (Cebesoy & Oztekin, 2018; Machluf & Yarden, 2013; Martins, Lencastre, et al., 2018). One teacher (5.56%) mentioned that "*The complexity of some processes and their interpretation by the students requires strong orientation and motivation, which should be taken into account when organizing the activity.*" This perspective was included in category: Constraints related with the student's performance.

Furthermore, the data showed that teachers considered that planning and implementing bioinformatics-based activities is more time-consuming and requires more resources than other activities (Q12.3.; Q12.4.) (Figure IV - 11). These notions can be

related with the lack of opportunities for teachers' training in bioinformatics. Training is crucial for teachers to feel more acquainted with bioinformatics tools and to clarify that planning and implementing bioinformatics-based activities can be framed within the time schedule for a class (90 minutes) as described by Martins, Fonseca and Tavares (2018). Another reason that can explain this result is the absence of didactic bioinformatics resources in Portuguese. In fact, the idiom of the majority of the platforms and of the exercises available is English which was previously reported as a barrier to non-English speakers (Machluf & Yarden, 2013; Martins, Lencastre, et al., 2017, 2018). This result highlights the importance of creating a portfolio of bioinformatics-based activities, in this case adapted to Portuguese, and making it available for the educational community in other to emphasize the adequacy of integrating bioinformatics in educational approaches.

5.3. Perceptions Regarding Workshop Attendance

A careful analysis of teachers training on bioinformatics was carried out having in consideration the previous results reported by Martins, Lencastre & Tavares (2017, 2018), by Machluf et al. (2012, 2013) and by Marques et al. (2014). Despite the interest of all the participants in attending training courses on bioinformatics for teachers (Q16), it was mentioned that the availability of these courses is still very scarce (Q12.5.) (Figure IV - 11). In fact, the importance of adequate teachers training in this scientific field, is further supported by the higher perception teachers have about their knowledge in bioinformatics after attending the workshop (Q4.2.), in comparison with that perception before the workshop (Q4.1.) (Figure IV - 11).

Regarding the workshop itself, teachers were asked about the reasons that motivate them to attend this workshop and to choose this workshop among others (Q13). They were asked to list the main difficulties found when they were performing the workshop activities (Q14) and suggestions were collected in how to improve the workshop (Q15).

The 18 participants listed the main reasons that motivated them to attend the workshop (Q13). By carrying out an analysis of the five most frequent words mentioned, it was possible to list: class (6.84%); application (5.98%); classroom (5.13%); learn (4.27%); and utilization (4.27%). This general analysis suggests that teachers wish to learn how to apply and how to use the bioinformatics tools in their classes. Adding to this analysis, it was possible to identify three main categories of answers. "*To learn to apply*" – was mentioned by three participants (16.67%). These participants revealed that they chose this workshop in order to learn "*how to implement the tools of these area*" in their classrooms. Six out of the 18 participants (33.33%) highlighted as the main reason to

attend this workshop "*Curiosity*". Teachers justified their attendance as an opportunity to learn more about bioinformatics, by curiosity and because they were "*interested in related topics such as genetics and all the fields that are related with DNA*". This result reinforces teachers' interest in this scientific topic. The third category is particularly important in the context of this study: "*The need of updating*". Mentioned by 9 participants (50.00%), this category included answers such as "*I felt the need of learning about this area* (…) *and to explore bioinformatics in a scientific and right perspective*" or "*I urgently need to improve my skills* (…) *to follow the quick development of these applications* (…) *and to implement them with my classes*". This result is in line with the considerations about the urgency of training courses reported above. Teachers are interest in this scientific area and the adequacy of the proposals to the schools is recognized, although they do not feel prepared to proceed with the implementation of the tools without previous specialized training (Machluf & Yarden, 2013; Martins, Lencastre, et al., 2018; Shuster, et al., 2016; Wood & Gebhardt, 2013).

In which concerns with the main difficulties found by the participants while performing the activities of the workshop (Q14), three categories of answers were defined. 50.00% of the participants (9 out of 18) listed as the main difficulties' technical aspects such as internet access, and the lack of time or difficulties to read the paper version of the guidelines. In fact, these constraints were related with the organization of the workshop itself and not with difficulties related with the bioinformatics-based activities performed. However, the internet connection was fixed even during the workshop and the digital version of the guidelines was sent by email to the participants. In this regard, we can assume that the logistic constraints were solved and the workshop workflow was not affected. Two out of 18 participants (11.11%) revealed to have difficulties in interpreting "*aspects associated with gene sequences/nucleotides*" and to "*understand the steps to follow*". In contrast with these results, 38.89% (7 out of 18) reported no difficulties and one teacher did not answer this question.

Finally, suggestions for workshop improvement were asked to the participants (Q15). 44.44% of the participants did not answer this question. 33.33% of the participants listed suggestions for improvement. Essentially participants claim for: "*More exercises*"; "*I would like to attend a training course (longer) in this area, once 3 hours are not sufficient to understand all the information discussed*"; and "*to increase the font size of the text in the guidelines – paper version*". 22.22% highlighted that "*do not think that the workshop needs to be changed*".

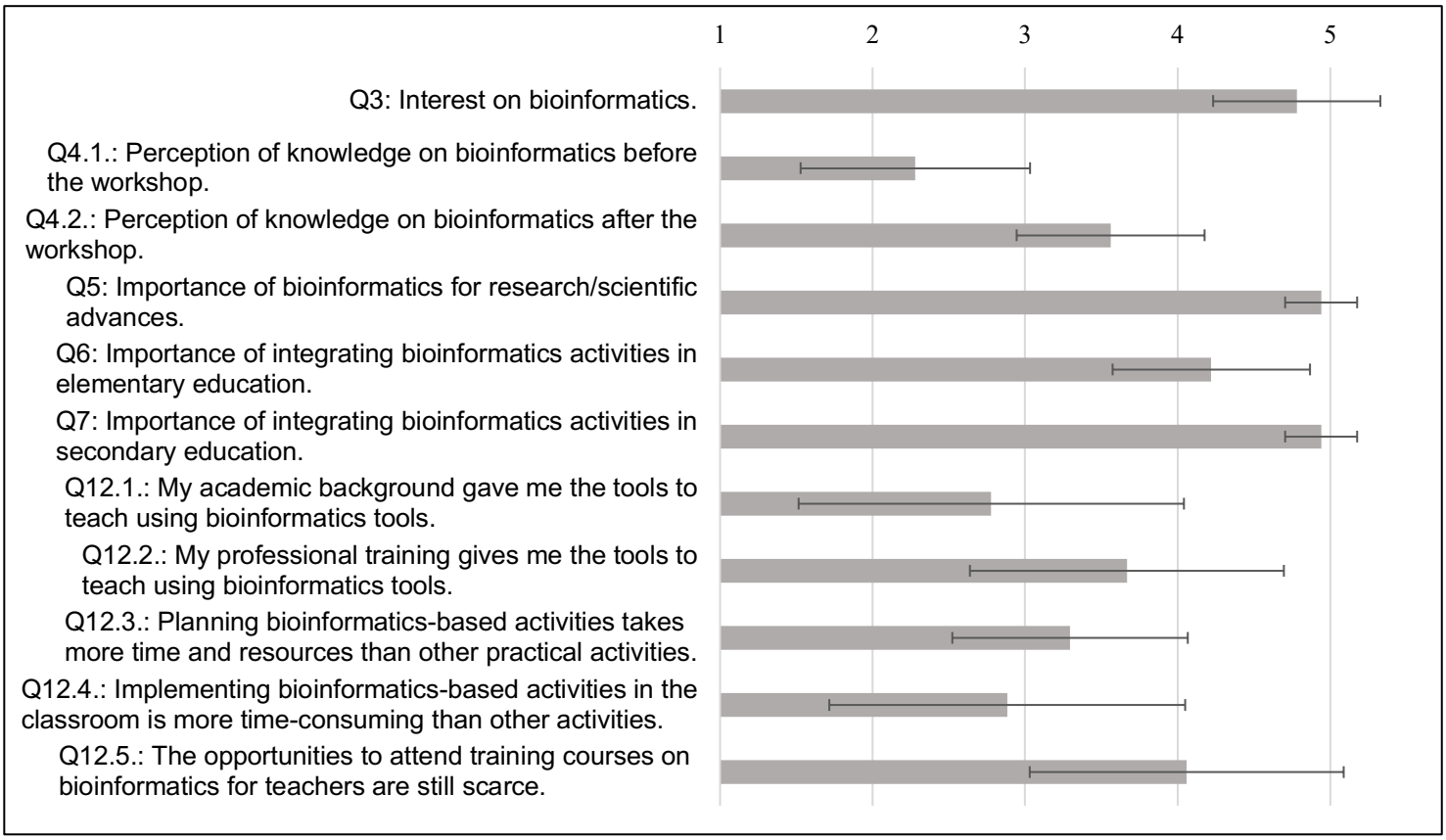


Figure IV-11. Answers given by participants according to a Likert Scale (Range 1 to 5). Grey bars represent the mean value and the error bars refer to the standard deviation.

6. Conclusion

Teachers revealed to be interested in bioinformatics and recognized its importance for scientific advances, which is in frame with the expected teachers' perceptions about bioinformatics as a scientific discipline. Teachers were open and motivated to integrate bioinformatics in their teaching practices. Regardless their will, teachers believe that key constraints have to be overcome, emphasizing the need of suitable training through dedicated courses. Thus, the main take-home message from this workshop is the urgent need of training courses for teachers in order to fuel the integration of bioinformatics in the curriculum and education daily practices. Adding to this, it is important to better understand the reasons why teachers admitted that their schools have the necessary conditions to implement bioinformatics-based approaches, but contradictorily they indicated as the main constraints to this implementation: poor internet connection and lack of computers. A reedition of the workshop occurred in July 2019 with 40 participant teachers. In this workshop, new data were collected in order to increase the robustness of this study. Adding to this, a website is under construction and will be soon available for teachers, with bioinformatics-based exercises in Portuguese in order to meet the participants request of having more available resources in their native language.

We believe that the current study is a wakeup call for educational stakeholders to boost bioinformatics educational integration aiming to meet the challenges of a society capable to understand the scientific advances and take informed decisions.

Acknowledgements

The authors are grateful to all participants of this study and to Casa das Ciências (https://www.casadasciencias.org/) for the invitation to present a workshop and for creating the opportunity to perform this research. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Attwood, T. K., Blackford, S., Brazas, M. D., Davies, A., & Schneider, M. V. (2017). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics*, *20*(2), 398-404. https://doi.org/10.1093/bib/bbx100

Casa das Ciências. (2018). *V Encontro Internacional da Casa das Ciências - Ciência, Comunicação, Imagem e Tecnologia*. https://www.casadasciencias.org/5encontrointernacional/

Cebesoy, U., & Oztekin, C. (2018). Genetics Literacy: Insights From Science Teachers' Knowledge, Attitude, and Teaching Perceptions. *International Journal of Science and Mathematics Education*, *16*(7), 1247–1268. https://doi.org/10.1007/s10763-017-9840-4

Chiovitti, A., Thorpe, F., Gorman, C., Cuxson, J., Robevska, G., Szwed, C., Duncan, J., Vanyai, H., Cross, J., Siemering, K., & Sumner, J. (2019). A citizen science model for implementing statewide educational DNA barcoding. *PLoS ONE*, *14*(1), e0208604. https://doi.org/10.1371/journal.pone.0208604

Eklund, J., Rogat, A., Alozie, N., & Krajcik, J. (2007). *Promoting Student Scientific Literacy of Molecular Genetics and Genomics*. http://www.umich.edu/~hiceweb/presentations/documents/Genetics_NARST_07.pdf

European Comission. (2013). *Survey of Schools: ICT in Education - Benchmarking Access, Use and Attitudes to Technology in Europe's Schools*. Publications Office of the European Union.

Foster, M., & Sharp, R. (2007). Share and share alike: deciding how to distribute the scientific and social benefits of genomic data. *Nature Reviews Genetics*, *8*(8), 633–639. https://doi.org/10.1038/nrg2124

Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research

simulation in bioinformatics. *Journal of Biological Education*, *40*(3), 107–112. https://doi.org/10.1080/00219266.2006.9656026

Kirkpatrick, G., Orvis, K., & Pittendrigh, B. (2002). A teaching model for biotechnology and genomics education. *Journal of Biological Education*, *37*(1), 31–35. https://doi.org/10.1080/00219266.2002.9655843

Koehler, M. J., & Mishra, P. (2000). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70. https://www.learntechlib.org/p/29544/

Kolarova, T. (2011). Modern biotechnology from the point of view of 15-19-year-old high school students. *Biotechnology and Biotechnological Equipment*, *25*(3), 2538–2546. https://doi.org/10.5504/bbeq.2011.0069

Kovarik, D., Patterson, D., Cohen, C., Sanders, E., Peterson, K., Porter, S., & Chowning, J. (2013). Bioinformatics Education in High School: Implications for Promoting Science, Technology, Engineering, and Mathematics Careers. *CBE—Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Lock, R. (1996). Biotechnology and genetic engineering: Student knowledge and attitudes: Implications for teaching controversial issues and the public understanding of science. In G. Welford Osborne, J. , Scott, P. (Ed.), *Research in Science Education in Europe* (pp. 229–242). The Falmer Press.

Luscombe, N., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, *40*(4), 346-358. PMID: 11552348

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible—the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Machluf, Y., Gelbart, H., & Yarden, A. (2012). High-School Teachers' Appropriation of an Innovative Curriculum in Bioinformatics. In D. Kruger & M. Ekborg (Eds.), *The 9th Conference of European Researchers in Didactics of Biology (ERIDOB)*.

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Mariano, D., Martins, P., Santos, L., & Minardi, R. (2019). Introducing Programming Skills for Life Science Students. *Biochemistry and Molecular Biology Education*, *47*(3),

288–295. https://doi.org/10.1002/bmb.21230

Marques, I., Almeida, P., Alves, R., Dias, M., Godinho, A., & Pereira-Leal, J. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r1_id=

Martins, A., Lencastre, L., & Tavares, F. (2018). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. In *3rd International Conference on Teacher Education (INCTE)*. Instituto Politécnico de Bragança. http://hdl.handle.net/10198/17381

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In M. Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

Mumtaz, S. (2000). Factors affecting teachers' use of information and communications technology: a review of the literature. *Journal of Information Technology for Teacher Education*, *9*(3), 319–342. https://doi.org/10.1080/14759390000200096

National Research Council. (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

Nunes, R., Júnior, E., Menezes, I., & Malafaia, G. (2015). Learning nucleic acids solving by bioinformatics problems. *Biochemistry and Molecular Biology Education*, *43*(5), 377–383. https://doi.org/10.1002/bmb.20886

Oliver, J., Slashinski, M., Wang, T., Kelly, P., Hilsenbeck, S., & McGuire, A. (2012). Balancing the Risks and Benefits of Genomic Data Sharing: Genome Research Participants' Perspectives. *Public Health Genomics*, *15*(2), 106–114. https://doi.org/10.1159/000334718

Punch, K. F. (2009). *Introduction to Research Methods in Education.* SAGE Publications.

Roberts, C. (2015). Content Analysis. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (pp. 769–773). Elsevier Inc. https://doi.org/10.1016/B978-0-08-097086-8.44010-9

Sadek, H. (2004). *Bioinformatics : principles, basic internet applications*. Trafford Publishing.

Sayres, M., Hauser, C., Sierk, M., Robic, S., Rosenwald, et al. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS ONE*, *13*(6), e0196878. https://doi.org/10.1371/journal.pone.0196878

Schuster, C., & Weber, R. (2006). Basic Content Analysis. *Journal of Marketing Research*, *23*(3), 310. https://doi.org/10.2307/3151496

Shuster, M., Claussen, K., Locke, M., & Glazewski, K. (2016). Bioinformatics in the K-8 Classroom: Designing Innovative Activites for Teacher Implementation. *International Journal of Designs for Learning*, *7*(1), 60–70. http://www.ncbi.nlm.nih.gov/pubmed/27429860

Tarver, T. (2010). Genomics: A New Challenge in Consumer Health Information Literacy. *Journal of Hospital Librarianship*, *10*(1), 23–32. https://doi.org/10.1080/15323260903470910

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School—New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Bioinformatics, a Befitting Tool for *e*-learning: Potential and Constraints according Teachers' Perceptions

Abstract. *Bioinformatics tools are suitable didactic instruments to combine updated knowledge with spotlight teaching strategies, e.g., e-learning. This study depicts the status of computational resources at schools pinpointed by teachers. Frequently, computers are obsolete, with outdated software, not connected to internet, their number is limited and often placed in areas not primarily aimed for teaching (schools' libraries and/or classrooms for professional/technical programs). These are key limitations preventing the implementation of digital-based activities in classrooms. This reality calls for the need to provide schools with updated informatics resources and fast internet connection to scaffold top-notch learning.*

Keywords. Bioinformatics, *e*-learning, Informatics at School, Teacher Professional Development

1. Introduction

Computational resources and biological research are strongly connected (Institute of Medicine and National Research Council, 2006; Iranbakhsh & Seyyedrezaei, 2011; Nature Cell Biology, 2000). Nowadays, every biology laboratory has computers with robust internet access to perform bioinformatics analysis (Huerta et al., 2000; Luscombe et al., 2001). In this regard and trying to adequate modern teaching practices to the new developments in biological research, several initiatives have been implemented to integrate bioinformatics-based approaches as an educational strategy (Machluf et al., 2017; Machluf & Yarden, 2013; Wefer & Sheppard, 2008; Wood & Gebhardt, 2013). In this scope, a portfolio of bioinformatics activities is available for the educational community to approach scientific issues such as antibiotic resistance, genetics, food preservation or evolution (Arnold et al., 2017; Bacusmo et al., 2019; Fernandes et al., 2014; Martins, Fonseca, et al., 2018).

Recognizing the key role of teachers in this process, research institutions, universities, policy makers and teachers training programs, need to focus their interventions in helping teachers to implement bioinformatics in their classes (Brown et al., 2014; Houseal et al., 2014; Machluf et al., 2012; Marques et al., 2014).

Several studies were carried out to diagnose teachers' perceptions about bioinformatics integration in middle and high school curricula (Machluf et al., 2012;

Martins et al., 2017; Martins, Lencastre, et al., 2018a; Martins et al., 2020). Teachers revealed to be interested in bioinformatics and recognized the importance of its incorporation in the curricula. However, from these studies was not clear the impact of the availability of computational resources at schools, namely computers and internet access. This study is a diagnostic of in-service teachers' perceptions about the accessibility of computational resources at schools to promote digital-based teaching and learning approaches in general, and to implement bioinformatics activities in particular.

1.1. Study Context

In a previous study, we diagnose teachers' perceptions about bioinformatics and identify the constraints for bioinformatics integration in middle and high schools (Martins et al., 2020). In fact, Martins *et al.* (2020) showed that teachers were interested in bioinformatics as a scientific area and as a didactic resource. Teachers revealed to be acquainted with bioinformatics definition, well-aware that computer and consistent internet access is required for data mining of biological datasets to retrieve meaningful information. Alongside with teachers' need of further training to boost their confidence to carry out bioinformatics-based interventions (Martins et al., 2020), it is urgent to revise middle and high school curricula to fuel an effective integration of bioinformatics in teaching practices, which is in line with other studies (Machluf et al., 2012, 2017; Machluf & Yarden, 2013; Marques et al., 2014; Martins et al., 2017; Martins, Lencastre, et al., 2018a).

When in-service teachers were asked about the informatics readiness of their schools to promote bioinformatics-based approaches in their teaching practices, the majority admitted that their schools have the necessary conditions to implement bioinformatics-based approaches. Despite this, participants pointed out as one of the main constraints the poor internet connection and lack of computers (Martins et al., 2020).

The importance of clarifying these testimonials was reinforced particularly taking into account the recent need to rapidly implement *e*-learning strategies as a consequence of the compulsory containment due to COVID-19 pandemic (Sintema, 2020; United Nations Educational, Scientific and Cultural Organization, 2020; Zhou et al., 2020).

To fully investigate the reason why teachers acknowledge that although schools have the resources (i.e. computers and internet) these are unsuitable to implement bioinformatics activities, we carried out an inquire to 37 Biology teachers who attended a bioinformatics workshop for science teachers (Casa das Ciências, 2019).

To tackle this, two research questions were raised:

- How well are Portuguese middle and high schools prepared to implement bioinformatics in the classroom?

- Is the informatics equipment available at schools, updated and accessible for all teachers and at any time, *i.e.*, do schools have the equipment available to teachers and ready to be used in the classroom?

2. Methods

2.1. Participants

This research focus on a group of teachers who attended a training workshop in bioinformatics "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the Classroom*". This four-hours workshop occurred in Lisboa in the context of an annual international meeting for teachers (Casa das Ciências, 2019).

This group consisted of 37 Biology teachers from 28 schools (26 public and 2 private) from 8 different regions, mainly urban areas, being 78.40% of the participants teaching at schools in Lisboa and Setúbal which are built-up areas with a high population density.

Eight of the 37 teachers hold a MSc degree. One teacher holds a PhD degree. Participants have an average of 25.44 ± 7.75 years of teaching experience. At the moment of the workshop, 12 teachers taught at middle school level (students between 12-15 years old), 12 taught at high school level (students between 16-18 years old) and 12 taught both middle and high school levels. One participant did not fill in this information.

2.2. Materials

2.2.1. The Workshop

The workshop "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the Classroom*" was designed and implemented for the first time in 2018 under the scope of the V International meeting for teachers of Casa das Ciências (Casa das Ciências, 2018). This workshop is aimed to explore with teachers the potential of bioinformatics as a didactic resource. Following specifically designed guidelines (Martins, Fonseca, et al., 2018), teachers are guided to explore four bioinformatics tools in order to data mining a DNA sequence focusing on identifying genes and determine the putative functions of their products. Additionally, using bioinformatics resources of comparative genomics, the presence of certain genes in different taxonomic groups is also analysed in order to infer evolutionary relationships. This holistic approach contributes to understand basic notions of genomics, genes, genomes and proteins and,

adding to this, introduces genomics-related key concepts such as Open Reading Frame (ORF), Basic Local Alignment Tool (BLAST) or synteny (Martins & Tavares, 2018).

In 2019, a reedition of the workshop was considered and performed during the VI International meeting for teachers of Casa das Ciências (Casa das Ciências, 2019). The workshop was updated to include further resources. Efficient Database framework for comparative Genome Analyses using BLAST score Ratios (EDGAR) platform was added in the workflow of the workshop to run genome comparisons (Justus-Liebig-Universitat, 2020). The activity was intended to identify, among up to five bacterial strains, the set of homologous and specific genes of each strain using the Venn Diagram functionality of EDGAR (Figure IV - 12). Based on the results obtained, notions of core genome, pan genome and accessory genome are discussed.

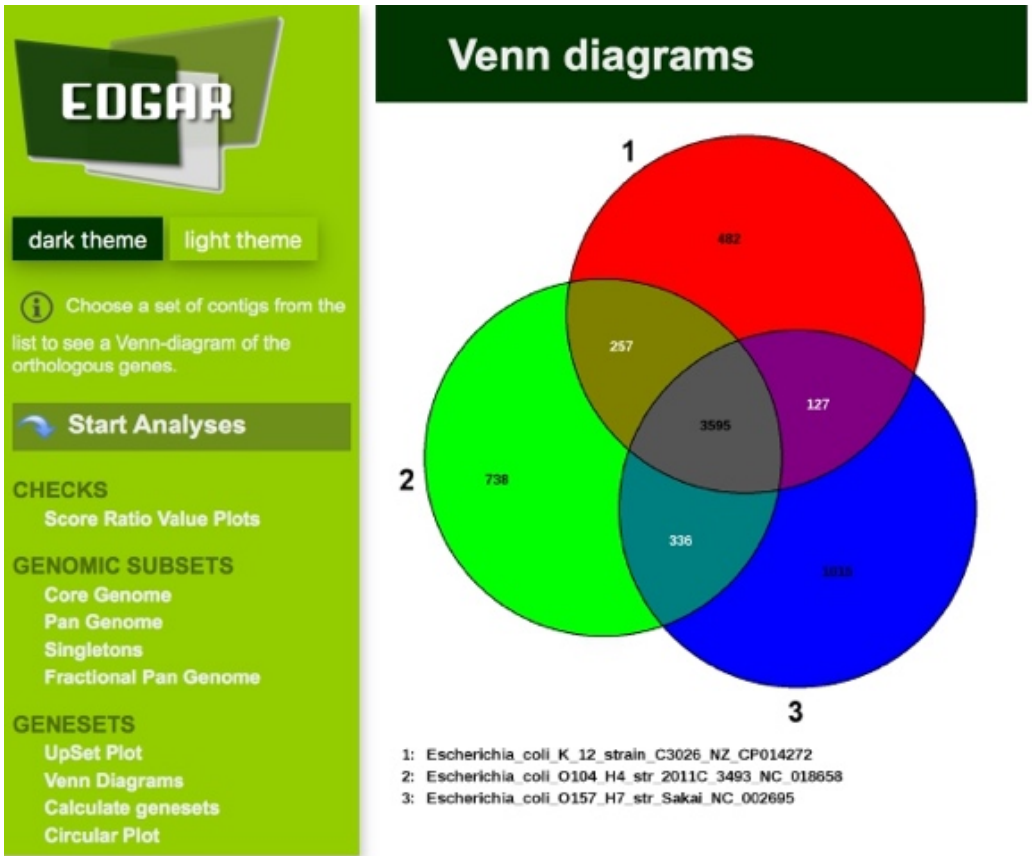


Figure IV-12. Genome comparison between three strains of *Escherichia coli* using a Venn Diagram to identify the core genome, pan genome and accessory genome.

Circular Plots and the Nucleotide Identity Average matrix (ANI) (Figure IV - 13) are analysed from a comparative genomics perspective. EDGAR functionality for the creation of phylogenetic trees is also explored.

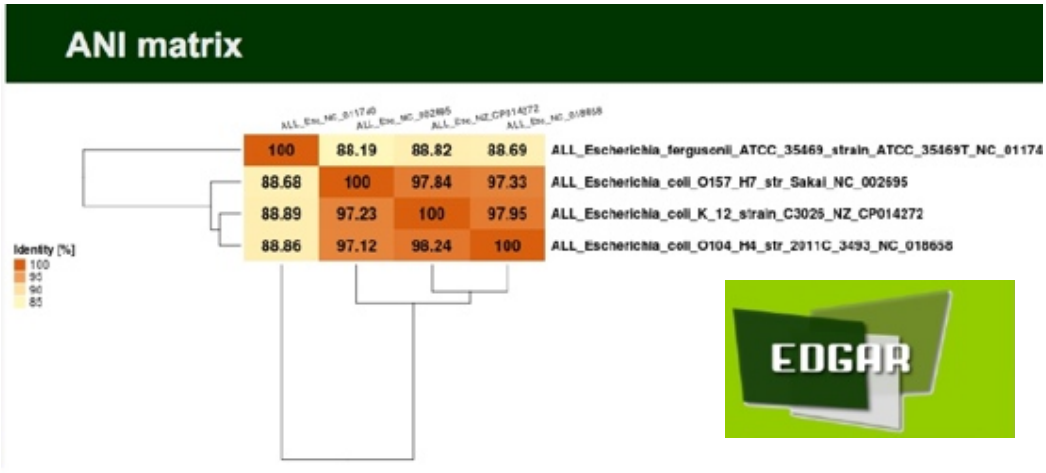


Figure IV-13. Genome comparison between a set of three strains of *Escherichia coli* and one strain of *Escherichia fergusonii* using an ANI matrix to identify genomes that belong to the same bacteria species (>95% according to Richter & Rosselló-Móra (2009)).

All the exercises proposed in the workshop privileged simple, intuitive and user-friendly tools. Adding to this, the graphic interface of the outputs obtained at EDGAR are appealing and empower analytical skills of data interpretation through graphs and/or tables.

2.2.2. The Questionnaire

The questionnaire (Figure IV - 14), developed in our previous study (Martins et al., 2020), was clustered in two dimensions: teachers perceived knowledge on bioinformatics (Q1, Q2, Q3, Q4, Q5, Q6, Q7, Q9, Q12, Q13, Q14, Q15, Q16) and in-service teachers' perceptions about the computational and internet resources available at their schools to implement bioinformatics-based activities (Q8, Q10, Q11).

These two dimensions aimed to know if the participants are acquainted with the definition and scope of bioinformatics, and to collect data on their perceptions of school readiness to implement bioinformatics as a didactic tool.

The questionnaire also included an initial section for demographic characterization of the group and three additional items to assess teachers' opinions about the questionnaire itself.

Teachers' perceived knowledge on bioinformatics

Q1: What is *Bioinformatics* for you?

Q2: Bioinformatics-based activities are more suitable to be framed: **(a)** in the Biology curricula; **(b)** in the Information and Communication Technology (ICT) curricula; or **(c)** in both Biology and ICT curricula.

Q3: Rate your interest on Bioinformatics: 1 (*Not interested at all*) - 5 (*Very interested*)

Q4: Rate your perception of knowledge on bioinformatics: **Q4.1:** Before the workshop: 1 (*Insufficient*) - 5 (*High*); **Q4.2:** After the workshop: 1 (*Insufficient*) - 5 (*High*)

Rate the importance … 1 (*Not important at all*) - 5 (*Very important*): **Q5:** … of bioinformatics for research and scientific advances; **Q6:** … of integrating bioinformatics activities in elementary education; **Q7:** … of integrating bioinformatics activities in secondary education.

Q9: Have you explored bioinformatics tools by yourself in order to implement bioinformatics-based activities in your classes? *Yes* __ *No* __

Q9.1: If so, did you implement the explored resources in the classroom? *Yes* __ *No* __

Q9.1.1: If not, please indicate the main reasons why you do not implement the resources in the classroom.

Q12: Please rate your agreement with the following sentences; 1 (*I totally disagree*) – 5 (*I totally agree*): **Q12.1:** My academic background gave me the tools to teach using bioinformatics tools; **Q12.2:** My professional training gives me the tools to teach using bioinformatics tools; **Q12.3:** Planning bioinformatics-based activities takes more time and resources than other practical activities; **Q12.4:** Implementing bioinformatics-based activities in the classroom is more time-consuming than other activities; **Q12.5:** The opportunities to attend training courses on bioinformatics for teachers are still scarce.

Q13: Indicate the main reasons that motivated you to attend this workshop.

Q14: List the main difficulties that you found while performing the activities proposed in this workshop.

Q15: Please make suggestion(s) for improvements that you consider important concerning the activities of the workshop you attended.

Q16: Would you be interested in attending more training courses/workshops promoted by research groups which use bioinformatics tools in their lab routines? *Yes* __ *No* __

Computational and internet resources at schools

Q8: Do you frequently use computers/tablets to explore online resources in practical classes? *Yes* __ *No* __

Q8.1: If so, please indicate how frequently in a school year do you use computers/tablets to explore online resources in practical classes: 1 (*Never*) – 5 (*Very often*)

Q8.2: If not, please indicate the main reason(s) why you do not frequently use computers/tablets to explore online resources in practical classes.

Q10: List the main constrains that can arise when implementing bioinformatics-based activities in the classroom.

Q11: Do you think that your school/institution has the needed conditions (computers and internet access) to explore bioinformatics in the classroom? *Yes* __ *No* __

Figure IV-14. Questionnaire used in the study.

2.3. Data Collection

Teachers voluntarily enrolled in the workshop which included a theoretical part and a practice component during which teachers co-worked in teams of two or three participants.

After the workshop all the participants were informed about the main aim of this study and, with their consent, the questionnaire (Figure IV - 14) was applied.

2.4. Data Analyses

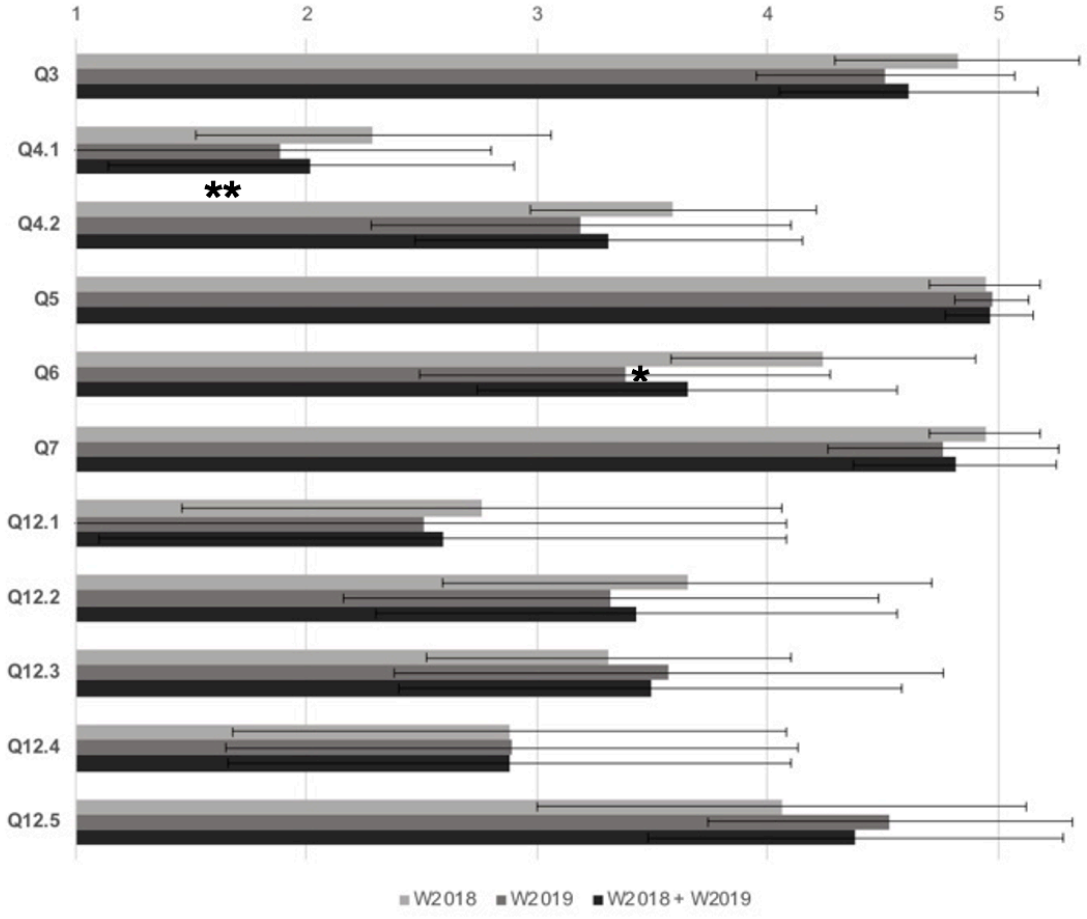
Descriptive and inferential statistical analysis were performed for quantitative data

(Punch, 2009). For qualitative data, a thematic content analysis of the participants' responses to open-ended questions was carried out (Roberts, 2015; Schuster & Weber, 2006).

3. Results and Discussion

3.1. Teachers' perceived knowledge on bioinformatics

The participant teachers of this study revealed to be aware of the main scope of bioinformatics field (Q1). The majority (54.29%) of the participants defined bioinformatics according to the etymology of the word, which is a one-dimensional definition referring to the application of information technology to biology. However, data analysis and data storage were also mentioned. In this regard, it can be considered that participants knew what bioinformatics is, recognized its importance for scientific research (Q5) and revealed to be highly interested in this scientific topic (Q3) (Figure IV - 15).



**Statistically significant differences between groups.*

***Statistically significant differences before and after the workshop.*

Figure IV-15. Answers given by participants according to a Likert Scale (Range 1 to 5). Bars represent the mean value and the error bars refer to the standard deviation.

Participants (51.35%) agreed that bioinformatics can be framed in Biology classes and understand its potential as a teaching and learning topic (Q2). Interestingly, teachers added that bioinformatics could also be explored both in Biology classes and Information and Communications Technology (ICT) classes. The reasoning is that bioinformatics can be framed in the Biology curricula, but it can also be used to promote interdisciplinary based pedagogies. These approaches are strongly encouraged according to up-to-date science teaching standards, such as the Next Generation Science Standards (National Research Council, 2013), especially in the Science, Technology, Engineering and Mathematics (STEM) field (Broggy et al., 2017; Cooper et al., 2017; National Research Council, 2012; You, 2017). This result reveals that participants of the study were aware of the adequacy of bioinformatics for Biology classes, but also went further and showed to understand the follow-up potentialities of this integration for other curricular areas. It is important to emphasize that participants were science teachers and consequently do not teach Information and Communications Technology (ICT) classes.

There was a general agreement among participants that bioinformatics-based activities are adequate for high school level (Q7) (Figure IV - 15). However, teachers inquired in this study, showed concerns about the importance of integrating bioinformatics-approaches in middle school level (Q6) (Figure IV - 15). This result is lower and statistically significant ($p<0.01$) when compared to the results obtained among the teachers who participated in the workshop edition of 2018 that more confidently agree on the importance of integrating bioinformatics in middle school (Martins et al., 2020). This difference can be due to the exercises explored in workshop edition of 2019 that broaden the range of platforms explored when compared with the first edition of the workshop. This could have contributed to teachers better understand how they integrate bioinformatics in high school classes and, at the same time, feel that all bioinformatics tools are too complex for middle school level. When designing a new workshop other platforms can be explored such as Pathogen Modeling Program (PMP) (United States - Department of Agriculture, 2020) or Combined Database for Predictive Microbiology – ComBase (Baranyi & Tamplin, 2004; University of Tasmania & United States - Department of Agriculture, 2020) that showed to be compatible with middle school level to explore, for example, food preservation techniques (Fernandes et al., 2014; Martins, Lencastre, et al., 2018b).

Not surprising, and consolidating the data obtained in the first study (Martins et al., 2020), the majority (70.27%) of participants admitted that they have never explored bioinformatics tools by themselves (Q9) and most of them (72.73%) revealed to have actually implemented the bioinformatics tools in their classes (Q9.1). Teachers confirmed

the perception that bioinformatics-based strategies are more time consuming and requires more resources than other type of practical classes (Q12.3) (Figure IV - 15).

Furthermore, they feel that their academic background and professional training is not sufficient for them to confidently explore bioinformatics tools within a didactic context (Q9.1.1; Q12.1; Q12.2) (Figure IV - 15).

Concerning the perceived background of teachers on bioinformatics, there is a statistically significant difference between teachers' answers ($p<0.01$) before and after the workshop (Q4.1; Q4.2) (Figure IV - 15). Teachers clearly agree that the workshop contributed to deeper their background on bioinformatics. In fact, workshop participants admitted that their background on bioinformatics improved after the workshop, boosting their confidence to explore bioinformatics in the classroom. Interestingly, the reason that motivated around half (48.65%) of the teachers to participate in the workshop (Q13) was to gain further training, corroborating the previous study (Martins et al., 2020).

These results were also obtained in the assessment of other training interventions on bioinformatics for teachers and corroborate the need of teachers update on this field (Q12.5.; Q16) (Machluf et al., 2017; Machluf & Yarden, 2013; Marques et al., 2014; Martins et al., 2017, 2020; Martins, Lencastre, et al., 2018a) (Figure IV - 15).

Regarding the workshop itself (Q14, Q15), most of the participants did not mention any improvements on the bioinformatics-based activities explored at the session. The ones who did, claimed for a longer workshop (more than 4 hours) or a 25 hours training course to broaden their perspectives on the potential of bioinformatics-based tools adapted to different school levels. Adding to this, informally teachers express their will to access scientific counselling to implement bioinformatics in their classes all over the school year.

3.2. Teachers' Perceptions about the Computational and Internet Resources Available at Schools to Implement Bioinformatics-Based Activities

According to the first dimension of questions, regarding teachers' perceived knowledge on bioinformatics, it is legitim to assume that participants were aware of the main aim of bioinformatics, of its potential as an educational resource, and that teachers were interested and motivated to learn more about this scientific field.

These participants were also conscious of what is needed in order to implement bioinformatics in their classes. Accordingly, they are able to have a critical and a helpful

perception about the possibilities and the constraints to integrate bioinformatics in their different school realities.

The second dimension of questions aimed to diagnose in-service teachers' perceptions about the computational and internet resources available at their schools for bioinformatics-based activities.

More than 90% of the participants admitted using computers/tablets to explore digital resources in their classes (Q8) which indicates that teachers are used to take advantage of technologies in their classes (Q8.1).

When asked specifically about the readiness of their institutions to develop bioinformatics activities, the majority (62.16%) of the participants assumed that the school/institutions where they were teaching did not have the necessary conditions to integrate bioinformatics-based strategies (Q11). Although this data may apparently contradict what teachers mentioned in the first edition of the workshop (Martins et al., 2020), the evidence gathered in the current study suggests that teachers understood that the existence of computers and internet does not ensure by itself the possibility to carry out bioinformatics exercises.

Other constraints pointed out by teachers impairing the implementation of bioinformatics activities in the classroom (Q10) are related with: logistic constraints (75%); training needs (8.33%); literacy (perceived knowledge and skills) (13.89%); and student's performance (2.78%).

Regarding the logistics constraints, teachers mainly mentioned that computers at school are obsolete, not easy to access and internet connection is often poor (Figure IV - 16). Around 10% of teachers reported informatics-based resources understood as computers and internet limitations.

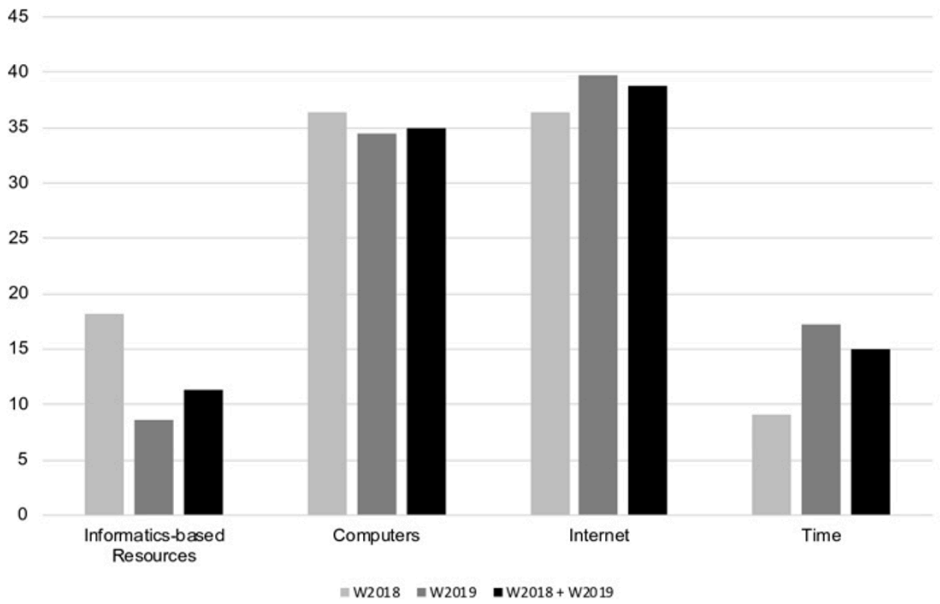


Figure IV-16. Main logistics constraints to implement bioinformatics in the classroom from an in-service teacher perspective (Q10).

Participant teachers who referred not to use technological equipment in their classrooms (Q8.2) mentioned reasons such as: "*Lack of computers and time-consuming administrative procedures to get access to a classroom equipped with computers*"; "*Difficulties to access the internet*" and "*Lack of computers in the schools and so teachers have to ask students to bring their personal computers*". These statements suggest that school computers are not really available for teaching, being frequently allocated to no teaching activities (school libraries and/or classrooms for professional/technical programs).

Furthermore, it is important to highlight that these results are likely biased since the inquired teachers serve in central schools located in large urban areas, in which is expected that informatics resources (i.e. computers and internet connection) are more likely accessible comparatively with non-urban schools. This suggests an inequity between urban and non-urban schools regarding the integration of bioinformatics in learning activities.

A possibility to overcome the lack of computers at school or to avoid using obsolete computers is inviting students to use their own. However, between 2009 and 2012, the percentage of students who reported having a least one computer or more at home in Portugal is lower than OECD average (Organisation for Economic Co-operation and Development, 2015). This reality become recently obvious with the *e*-learning strategies implemented due to COVID-19 pandemic (Kotowicz, 2020; Silva, 2020).

Computers are now a key tool for teaching and learning and more than ever their role as a didactic instrument, that can connect students and teachers, is highlighted (Basilaia & Kvavadze, 2020; Huang et al., 2020; Ting et al., 2020). In this regard, governments

should develop programs and create funding opportunities in order to make possible for each student to have a computer at home.

Alternatively, personal smartphones may be used to perform simple and accessible tasks that do not require a computer as for instance to introduce Phyton to answer biological questions (Rueda et al., 2019), or explore biodiversity using deep-learning platforms such as the iNaturalist® (INaturalist, n.d.; Unger et al., 2020).

Regarding limitations related with internet access, teachers stated: "*It is present at school, but it is not working in an efficient way*". In fact, among the schools sampled, although internet connection is available, its efficiency can only provide basic tasks, such as email or to access digital resources for teachers, thus not suitable for bioinformatics analysis. In public schools, the internet network is provided by the ministry of education and it has a limited access (both concerning speed and number of computers connected with). This means that even if the students bring their personal computers, a request for access a robust wireless connection has to be made to execute bioinformatics exercises. In this context, improving internet access to both teachers and students within the schools needs to be urgently considered by educational stakeholders.

Finally, it is worthy mention that technical support is important to ensure that informatics equipment is set to operate normally. Although this aspect was not mentioned by teachers, it was inferred from informal discussions at the workshop "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the Classroom*" (2019). Furthermore, interdisciplinarity and collaboration between Biology teachers and ICT teachers during the activities could help to address problems related with computers and internet connection.

4. Conclusion

Generally, teachers acknowledged that their schools are equipped with computers and internet connection (Martins et al., 2020). This may suggest that resources would be available to integrate bioinformatics in teaching practices. However, in the present study teachers admitted that often computers are obsolete with outdated software, poor internet connection and inaccessible for teaching.

Focusing on these considerations the active use of educational web-based resources, in which bioinformatics can have a key role, calls for a digital reform of schools as encouraged by Next Generation Science Standards (NGSS) (National Research Council, 2013).

Acknowledgements

The authors are thankful to all participants of this study and to Casa das Ciências (https://www.casadasciencias.org/) for the call to reedit the workshop, opening the possibility to develop this research. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Arnold, M., Holman, D., & Zweifel, S. (2017). Using Molecular Biology and Bioinformatics to Investigate the Prevalence of Mislabeled Fish Samples. *The American Biology Teacher*, *79*(9), 763–768. https://doi.org/10.1525/abt.2017.79.9.763

Bacusmo, J., Bokor, J., Savage, K., & Crécy-Lagard, V. (2019). Identifying Pathogenic Islands through Genome Comparison. *The American Biology Teacher*, *81*(8), 577–581. https://doi.org/10.1525/abt.2019.81.8.577

Baranyi, J., & Tamplin, M. (2004). ComBase: A common database on microbial responses to food environments. *Journal of Food Protection*, *67*(9), 1967–1971. https://doi.org/10.4315/0362-028X-67.9.1967

Basilaia, G., & Kvavadze, D. (2020). Transition to Online Education in Schools during a SARS-CoV-2 Coronavirus (COVID-19) Pandemic in Georgia. *Pedagogical Research*, *5*(4), em0060. https://doi.org/10.29333/PR/7937

Broggy, J., O'reilly, J., & Erduran, S. (2017). Interdisciplinarity and Science Education. In *Science Education* (pp. 81–90). SensePublishers. https://doi.org/10.1007/978-94-6300-749-8_6

Brown, J., Bokor, J., Crippen, K., & Koroly, M. (2014). Translating Current Science into Materials for High School via a Scientist-Teacher Partnership. *Journal of Science Teacher Education*, *25*(3), 239–262. https://doi.org/10.1007/s10972-013-9371-y

Casa das Ciências. (2018). *V Encontro Internacional da Casa das Ciências - Ciência, Comunicação, Imagem e Tecnologia*. https://www.casadasciencias.org/5encontrointernacional/

Casa das Ciências. (2019). *VI Encontro Internacional da Casa das Ciências - Ensino das Ciências e a Sociedade Moderna*. https://www.casadasciencias.org/6encontrointernacional/

Cooper, K., McGraw, A., & Khazanchi, D. (2017). Bioinformatics for middle school aged children: Activities for exposure to an interdisciplinary field. *ISEC 2017 - Proceedings of the 7th IEEE Integrated STEM Education Conference*, 1–9.

https://doi.org/10.1109/ISECon.2017.7910217

Fernandes, E., Dias, C., Fonseca, M., & Tavares, F. (2014). Understanding Growth and Thermal Inactivation of Foodborne Bacteria Using the Pathogen Modelling Program (PMP). In Manuel Costa, P. Pombo, & J. Dorrio (Eds.), *Hands-on Science: Science Education with and for Society* (pp. 207–210). Hands-on Science Network.

Houseal, A., Abd-El-Khalick, F., & Destefano, L. (2014). Impact of a student-teacher-scientist partnership on students' and teachers' content knowledge, attitudes toward science, and pedagogical practices. *Journal of Research in Science Teaching*, *51*(1), 84–115. https://doi.org/10.1002/tea.21126

Huang, R., Liu, D., Tlili, A., Yang, & Wang, J. F. (2020). *Handbook on Facilitating Flexible Learning During Educational Disruption: The Chinese Experience in Maintaining Undisrupted Learning in COVID-19 Outbreak*. Smart Learning Institute of Beijing Normal University.

Huerta, M., Downing, G., Haseltine, F., Seto, B., & Liu, Y. (2000). *NIH Working Definition of Bioinformatics and Computational Biology*. https://2digitstechcom.ipage.com/uploads/2/9/0/1/2901227/compubiodef.pdf

INaturalist. (n.d.). *About · iNaturalist*. Retrieved May 17, 2020, from https://www.inaturalist.org/pages/about

Institute of Medicine and National Research Council. (2006). Globalization, biosecurity, and the future of the life sciences. In *Globalization, Biosecurity, and the Future of the Life Sciences*. National Academies Press. https://doi.org/10.17226/11567

Iranbakhsh, A., & Seyyedrezaei, S. (2011). The impact of information technology in biological sciences. *Procedia Computer Science*, *3*, 913–916. https://doi.org/10.1016/j.procs.2010.12.149

Justus-Liebig-Universitat. (2020). *EDGAR - "Efficient Database framework for comparative Genome Analyses using BLAST score Ratios."* https://edgar.computational.bio.uni-giessen.de/cgi-bin/edgar_login.cgi

Kotowicz, A. (2020). Alunos sem computador são mais do que se pensa. *O Observador*. https://observador.pt/2020/04/15/alunos-sem-computador-sao-mais-do-que-se-pensa-nas-escolas-publicas-quase-um-terco-dos-alunos-do-ensino-basico-nao-tem-equipamento/

Luscombe, N., Greenbaum, D., & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, *40*(4), 346–358. http://www.ncbi.nlm.nih.gov/pubmed/11552348

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible—the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Machluf, Y., Gelbart, H., & Yarden, A. (2012). High-School Teachers' Appropriation of an Innovative Curriculum in Bioinformatics. In D. Kruger & M. Ekborg (Eds.), *The 9th Conference of European Researchers in Didactics of Biology (ERIDOB)*.

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Marques, I., Almeida, P., Alves, R., Dias, M. J., Godinho, A., & Pereira-Leal, J. B. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r1_id=

Martins, A., Lencastre, L., & Tavares, F. (2018a). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. In *3rd International Conference on Teacher Education (INCTE)*. Instituto Politécnico de Bragança. http://hdl.handle.net/10198/17381

Martins, A., Lencastre, L., & Tavares, F. (2018b). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 309–317). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2020). Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers. *The Beauty and Pleasure of Understanding: Engaging with Contemporary Challenges Through Science Education*. European Science Education Research Association. *In Press.*

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In Manuel Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

National Research Council (2012). A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. In *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press. https://doi.org/10.17226/13165

National Research Council. (2013). *Next Generation Science Standards*. National Academies Press. https://doi.org/10.17226/18290

Nature Cell Biology. (2000). The importance of technological advances. *Nature Cell Biology*, *2*(3), E37–E37. Nature Publishing Group. https://doi.org/10.1038/35004064

Organisation for Economic Co-operation and Development. (2015). *Students, Computers and Learning: Making the connection*. https://doi.org/10.1787/9789264239555-en

Punch, K. (2009). *Introduction to Research Methods in Education.* SAGE Publications.

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Roberts, C. (2015). Content Analysis. In *International Encyclopedia of the Social & Behavioral Sciences: Second Edition* (769–773). Elsevier Inc. https://doi.org/10.1016/B978-0-08-097086-8.44010-9

Rueda, A., Benítez, G., Marchetti, J., Hasenahuer, M., Fornasari, M., Palopoli, N., & Parisi, G. (2019). Bioinformatics calls the school: Use of smartphones to introduce Python for bioinformatics in high schools. *PLoS Computational Biology*, *15*(2), e1006473. https://doi.org/10.1371/journal.pcbi.1006473

Schuster, C., & Weber, R. (2006). Basic Content Analysis. *Journal of Marketing Research*, *23*(3), 310. https://doi.org/10.2307/3151496

Silva, S. (2020). Autarquias investem milhões em computadores e Internet para alunos. *Público*. https://www.publico.pt/2020/04/24/sociedade/noticia/autarquias-investem-milhoes-computadores-internet-alunos-1913530

Sintema, E. (2020). Effect of COVID-19 on the Performance of Grade 12 Students: Implications for STEM Education. *EURASIA Journal of Mathematics, Science and Technology Education*, *16*(7), em1851.

Ting, D., Carin, L., Dzau, V., & Wong, T. (2020). Digital technology and COVID-19. *Nature Medicine*, *26*(4), 459–461. https://doi.org/10.1038/s41591-020-0824-5

Unger, S., Rollins, M., Tietz, A., & Dumais, H. (2020). iNaturalist as an engaging tool for identifying organisms in outdoor activities. *Journal of Biological Education*, 1–11. https://doi.org/10.1080/00219266.2020.1739114

United Nations Educational, Scientific and Cultural Organization. (2020). *UNESCO's support: Educational response to COVID-19*. UNESCO Building Peace in the Minds of Men and Women. https://en.unesco.org/covid19/educationresponse/support

United States - Department of Agriculture. (2020). *Pathogen Modeling Program (PMP) Online: Home*. https://pmp.errc.ars.usda.gov/default.aspx

University of Tasmania, & United States - Department of Agriculture. (2020). *ComBase - Combined Database for Predictive Microbiology*. https://www.combase.cc/index.php/en/

Wefer, S., & Sheppard, K. (2008). Bioinformatics in high school biology curricula: A study of state science standards. *CBE Life Sciences Education*, *7*(1), 155–162. https://doi.org/10.1187/cbe.07-05-0026

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School—New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

You, H. (2017). Why Teach Science with an Interdisciplinary Approach: History, Trends, and Conceptual Frameworks. *Journal of Education and Learning*, *6*(4), 66-77. https://doi.org/10.5539/jel.v6n4p66

Zhou, L., Wu, S., Zhou, M., & Li, F. (2020). 'School's Out, But Class' On', The Largest Online Education in the World Today: Taking China's Practical Exploration During The COVID-19 Epidemic Prevention and Control As an Example. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3555520

Chapter V

Assisting Teachers and Promoting Network

Chapter V includes the following publication:

Martins, A., Lencastre, L., & Tavares, F. (2020). "Bioinformática na Sala de Aula": Webpage to Boost Bioinformatics in the Classroom. *Simpósio Internacional de Psicologia Da Educação: Passado, Presente e Futuro* (SInPE20). *In Press.*

"Bioinformática na Sala de Aula": Webpage to Boost Bioinformatics in the Classroom

Abstract. *Bioinformatics has already proved to be a powerful didactic resource. Biology teachers' self-reflections about the reasons preventing them to integrate bioinformatics in educational practices reveal: the need of training, the lack of time and the scarce offer of resources in Portuguese. This study describes the design of the webpage "Bioinformática na Sala de Aula" aimed to scaffold teachers to integrate bioinformatics in their practices by meeting their reported needs. The webpage was designed taking into account the main constraints identified by teachers and focusing on scientific topics in which students' misconceptions were previously diagnosed. In this regard, activities' guidelines for teachers and students, Power Point presentations for teachers, supplementary information, tips and a discussion forum were developed and are available. The webpage has been continuously updated, in order to answer teachers' requests, and spread by partners such as Casa das Ciências. "Bioinformática na Sala de Aula" is a portfolio of resources validated as suitable didactic instruments and ready to be used in the classroom, saving teachers' time – activities handouts are in Portuguese and the topics approached are framed according to the Portuguese science curricula. The webpage is also a repositorium of genomics and bioinformatics-related information for teachers, being a complete source of information. Communication channels between experts on bioinformatics for schools and teachers are also provided in the webpage for teachers to feel continuously supported. In the future, we expect to address the impact that the webpage has been having in promoting bioinformatics-based learning.*

Keywords: Bioinformatics, Curricula, Self-reflection, Training, Webpage

1. Introduction

Bioinformatics has already proved to be a powerful didactic resource which combines an interdisciplinary approach with top-notch scientific research tools (Bloom, 2001; Kovarik et al., 2013; Lewitter & Bourne, 2011; Machluf & Yarden, 2013). In order to boost its integration in middle and high school level, a cooperative path between teachers and research institutions needs to be sharped (Machluf et al., 2017; Marques et al., 2014; Wood & Gebhardt, 2013). Teachers' training courses are seen as efficient initiatives to foster this networking and an opportunity to depict the main constraints that are preventing teachers from taking full advantage of bioinformatics tools in their teaching

practices. International studies have shown that teachers generally claim for training on bioinformatics-based activities curricular framed and for a scientific and technical continuous support. The lack of time available to perform *in silico* activities and language barriers are also underlined by teachers as major constraints that cannot be dismissed when trying to update teaching practices based on bioinformatics approaches (Kovarik et al., 2013; Machluf et al., 2017; Machluf & Yarden, 2013; Wood & Gebhardt, 2013).

In the Portuguese context, biology teachers' self-reflections further sustain the need for bioinformatics-dedicated training courses, the lack of time and the scarce offer of resources in Portuguese language (Marques et al., 2014; Martins, Lencastre, et al., 2020, 2018).

The diagnostic of the constraints that are preventing teachers from implementing bioinformatics-based approaches in their classes shed some light on the interventions that have to be designed. Among the strategies that have been proposed, are the creation of repositories of *in silico*-based activities focused on core competencies in bioinformatics aligned with schools' science curricula (Form & Lewitter, 2011). There are several repositories available online and open source, targeting different learning levels (middle and high school level, undergraduate level) with an offer of teaching resources that can be easily implemented in classrooms, without being an extra burden of work for teachers (European Learning Laboratory for the Life Sciences, 2020; Microbial Life, 2020; Network for Integrating Bioinformatics into Life Sciences Education, 2020). Furthermore, these databases of bioinformatics-based didactics resources are within webpages complemented with glossaries of core concepts, and forums to promote teachers' networking and to expand the interaction among teaching communities (Machluf & Yarden, 2013; Wood & Gebhardt, 2013; Zhang et al., 2007).

Therefore, webpages are extremely useful solutions to centralize the information that educators need and to facilitate networking between teachers and researchers, independently of their geographic location (Barker, 2009; Hakverdi-Can & Dana, 2012; Perrault, 2007). In this scope, the design of the webpage "*Bioinformática na Sala de Aula*" appears as a corollary of an effort to scaffold Portuguese science teachers to integrate bioinformatics in their practices, by meeting their reported needs, and to promote the use of open access bioinformatics resources with intuitive interfaces, inviting teachers to explore innovative and motivating pedagogical practices in formal and non-formal teaching contexts.

2. Methodology

2.1. The Webpage Design

The webpage (https://bioinformaticaaula.wixsite.com/bioinformatica-pt) was designed taking into account the main constraints to integrate bioinformatics in the classroom identified by teachers, namely time, language and the need for more training to improve their background on bioinformatics (Martins, Lencastre, et al., 2020).

To avoid time-consuming procedures to prepare and implement bioinformatics-based activities in the classroom, a portfolio of dry lab resources was compiled and made available in the webpage, organized according four main themes: *Molecular biology: in silico analysis* (Theme 1); Lac operon: gene regulation and evolutionary relationships (Theme 2); *Exploring metabolic pathways across the different life domains* (Theme 3); and *Bioinformatics in service to the population – practical examples* (Theme 4). All the resources were validated by in-service science teachers and, part of them, implemented in both formal and non-formal educational contexts in order to validate their impact on students' scientific and digital literacy, interest and attitudes (Martins, Fonseca, et al., 2020, 2018; Martins, Lencastre, et al., 2018). There are three main reasons behind the scientific topics chosen: firstly, they are related with students' pre-existing knowledge; secondly, the topics chosen drive to a meaningful introduction of new core concepts, enhancing students' literacy; and thirdly, the four thematic issues contribute to overcome students' misconceptions, namely the use of gene and genome as synonyms or misinterpretations regarding gene regulation (Form & Lewitter, 2011; Lewis & Kattmann, 2004; Martins, Fonseca, et al., 2020; Martins & Tavares, 2018; Shaw et al., 2008).

The webpage, written in Portuguese, consists of an Welcome Page (Página Inicial); Exercises (Exercícios) supported by guidelines and Power Point presentations; Supplementary Materials for Teachers (Materiais de Apoio ao Professor); a dedicated section for Training Courses' information (Cursos de Formação); a Forum (Fórum); details regarding School Partnerships (Escolas Parceiras) and About the team (Sobre Nós) (Figure V – 1). The webpage has been continuously updated in order to answer teachers' requests, and spread by partners such as Casa das Ciências and CIBIO – Research Centre in Biodiversity and Genetic Resources - InBIO Associate Laboratory through their webpages (Casa das Ciências, 2017; CIBIO-InBIO, 2020).

3. Webpage Contents

3.1. Welcome Page (Página Inicial)

In the welcome webpage (Figure V-1) the reader finds a brief description of the potential of bioinformatics as a didactic tool, supported by relevant state-of-the-art literature, and an explanation about the rationale behind the creation of the webpage "*Bioinformática na Sala de Aula*".

Figure V-1. Layout of the Welcome Webpage.

3.2. Exercises (Exercícios)

Exercises section compiles a portfolio of bioinformatics labs framed in the curricula, with detailed guidelines and Power Point presentations for their implementation, organized on four main topics (Theme 1 – Theme 4) (Figure V-2).

Figure V-2. Overview of Exercises section.

The rational to select the topics was grounded on their relevance according to the science curriculum and also having into consideration their potential to dismiss students' specific misconceptions. In this regard, *Molecular biology: in silico analysis* (Theme 1)

uses the bioinformatics platform "*In silico simulation of molecular biology experiments*" (Bikandi et al., 2004; Millán et al., 2013), to perform analysis of restriction enzymes, as well as to simulate *in silico* experiments of Polymerase Chain Reaction (PCR). Although the potential of this platform is broader, emphasis is given to two applications that meet the requirements of the school curricula: i) Amplification by PCR to reconstitute and confirm *in silico* a diagnostic experience of pathogenic bacteria reported in a scientific article; and ii) The elaboration of restriction maps of bacteria genomes using different restriction enzymes. Theme 1 exercises are also dedicated to explore DNA sequence analysis tools of the National Center for Biotechnology Information (NCBI) to identify putative genes (Martins, Fonseca, et al., 2018; NCBI Resource Coordinators, 2018).

Theme 2 – *Lac operon: gene regulation and evolutionary relationships* – was set up to acknowledge *lac* operon, an example frequently displayed in textbooks to approach gene regulation, and aware of its potential for comparative genomics studies. Framed within this theme, NCBI ORFfinder tool is used to identify start codons, namely alternative codons, and also to explore specificities of the bacterial genetic code (Martins, Fonseca, et al., 2018; National Center for Biotechnology Information, 2020). Microbial Genome Annotation & Analysis Platform (MaGe) is used to make evolutionary assumptions through comparative genomics (Vallenet et al., 2013). These exercises have been shown to benefit students learning regarding genomics-related concepts, as well as to update the currently taught notions (Martins, Fonseca, et al., 2020; Martins & Tavares, 2018).

In order to approach metabolic pathways using bioinformatics tools, Krebs cycle was the topic chosen for theme 3 – *Exploring metabolic pathways across the different life domains* – having in mind that it is a core concept of science school curricula. Within this theme, the exploration of the MetaCyc - Metabolic Pathway Database tool allows to study Krebs cycle and gives a valuable contribution to understand the amphibolic character of its intermediate compounds involved in other reactions of cellular metabolism (Caspi et al., 2017). In addition, in theme 3 an exercise is dedicated to study a biomolecule present in human metabolism – cholesterol, using HumanCyc - Encyclopedia of Human Genes and Metabolism (Romero et al., 2005). This platform makes possible to analyze in detail the metabolic pathway that gives rise to it, while recognizing other associated compounds that assume a crucial role in the human body (example: in the production of vitamin D). Furthermore, students' awareness of the connections between the biosynthesis pathways of different biomolecules, besides its scientific value, may be an endorser for the adoption of healthier behaviors.

Within Theme 4 – *Bioinformatics in service to the population – practical examples* (Figure V-3), bioinformatics labs were optimized to stress the importance of predictive bacteria growth models in food matrices, which is a subject particularly suitable for both middle and high school students, that when combined with wet lab procedures, can positively impact students' literacy on food preservation techniques and enhance their motivation as previously showed (Martins, Lencastre, et al., 2018).

At the bottom of the webpage "*Bioinformática na Sala de Aula*" (https://bioinformaticaaula.wixsite.com/bioinformatica-pt), there are several documents with tips and suggestions for teachers to implement the activities in their classes, such as the clarification that all the platforms and tools proposed to use are open access and the suggestion that some platforms as NCBI can be faster accessed during the morning in Portugal, mainly due to the time zone difference with countries where the number of researchers connected increase considerably during the afternoon. The link to the forum and to the main contact is also described in this section of the webpage, aiming to set an easy-going way for teachers to ask for support.

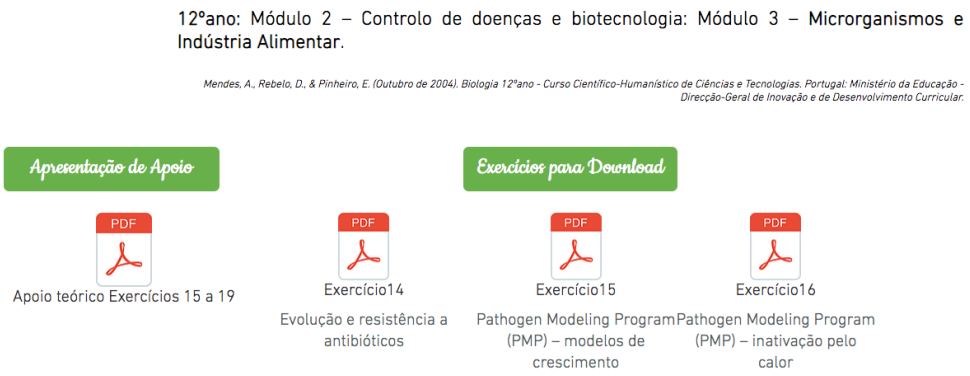


Figure V-3. Resources for teachers to address theme 4.

3.3. Supplementary Materials for Teachers (Materiais de Apoio ao Professor)

This section provides information that will assist teachers in implementing bioinformatics exercises, namely bibliographic references or information related to the contents approached in the bioinformatics labs such as genetic code, synteny or Open Reading Frames (ORF) (Figure V-4). This section is continuously updated in order to answer and clarify any questions that may arise, and that may be shared on the forum.

Figure V-4. Extra documents and useful online resources for teachers are provided to explain genomics core concepts.

3.4. Training Courses (Cursos de Formação)

A dedicated area for training courses is reserved to the dissemination of information. In the example below (Figure V-5), a section with exclusive access for the participants of the training course was created in order to share the guidelines of the course, as well as the projects produced by teachers, namely the training course assessment document. The sharing of these documents was provided after teachers' approval.

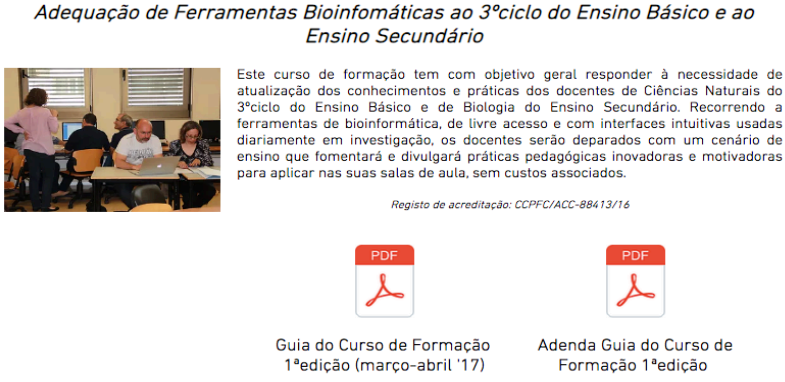


Figure V-5. The webpage is also a channel of dissemination of bioinformatics training courses for teachers.

3.5. Forum (Fórum)

A forum (Figure V-6) for sharing experiences, opinions, questions and comments is a privileged communication channel to boost collaborative work, encourage discussion of ideas and scaffold teachers.

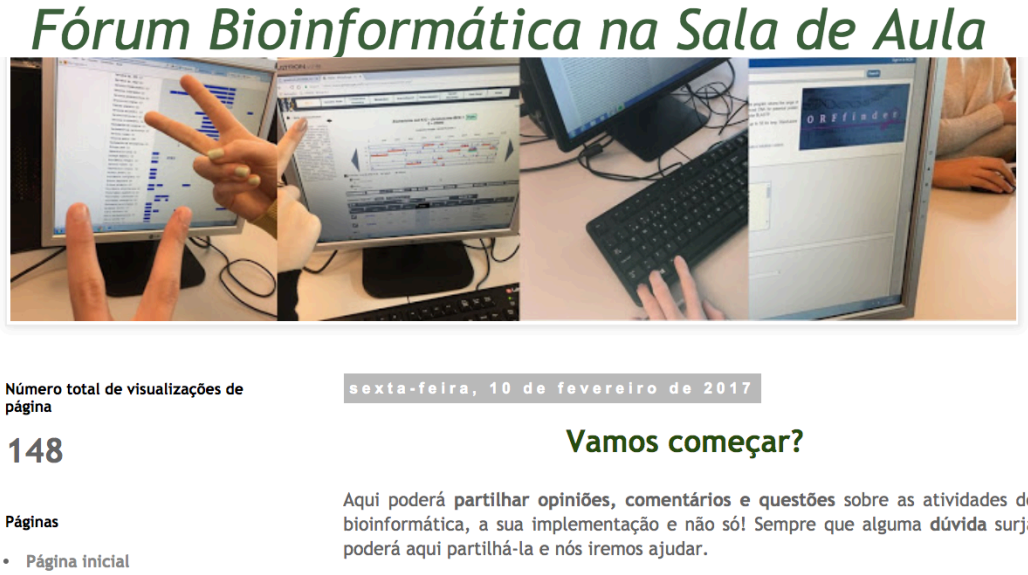


Figure V-6. Forum interface.

3.6. School Partnerships (Escolas Parceiras)

Partnerships section includes information regarding the schools that collaborated with the research group namely by allowing the implementation of some activities, such as "*Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes*" (Martins, Fonseca, et al., 2018). In order to publish school logos (Figure V-7), an authorization was obtained from each school directive board.

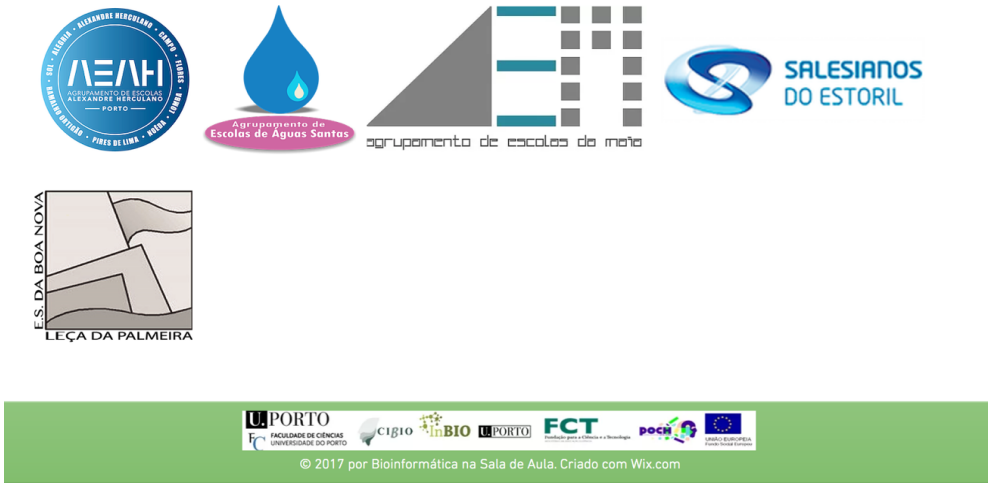


Figure V-7. School partnerships.

3.7. About (Sobre Nós)

In this section it is possible to obtain direct contact with the research group, as well as more detailed information about the team and the research aims of Microbial Diversity and Evolution (MDE) group from CIBIO-InBIO and hosted at Faculdade de Ciências da Universidade do Porto (FCUP) (Figure V-8).

Figure V-8. Details for contact are provided to the visitors.

4. Discussion

As previously mentioned, this webpage is focused on meeting Portuguese teachers' needs to implement bioinformatics in their schools. Despite the guidelines of the bioinformatics labs have been addressed in schools and validated for in-service teachers, no formal assessment of the use of this webpage by teachers was performed yet. Nevertheless, some considerations can be drawn such as the accesses to the webpage. It could be noticed that between 2017 and 2020, the access was biased by teachers who attended bioinformatics training and who had direct contact with the team of the project. Thus, the accession of this tool was not so frequent as desired, highlighting the need to disseminate the webpage using more proficient strategies. Designing an *e*-learning training course in which teachers have to access the webpage to accomplish specific tasks such as download the handouts, contribute with information as "*wikis*" or podcasts or to use the forum to register comments and doubts as a blog, should be thought as webpage promotion strategies (Boulos et al., 2006; Richardson, 2010; Wheeler & Wheeler, 2009).

Linked with the need to improve dissemination strategies, is the design of a research plan to evaluate the impact of the webpage on teaching practices. In the future, web-

based portfolio entries analysis (Oner & Adadan, 2011), online discussions on a private listserv complemented with attitude surveys (Koszalka, 2001), questionnaires based on models of technology acceptance (Akpinar & Bayramoglu, 2008; Liaw, 2002), that can also be online (Yuen et al., 2011), are reported strategies that can be taken into consideration to effectively depict the webpage impact and to overcome the constraints pointed out by teachers in promoting bioinformatics-based learning.

5. Conclusion

"*Bioinformática na Sala de Aula*" is a portfolio of resources validated as suitable didactic instruments and ready to be used by teachers in the classroom. Activities handouts are in Portuguese and the topics approached are framed according to the Portuguese science curricula, focus on contents particularly addressing students' difficulties or misconceptions. The webpage is also a repositorium of genomics and bioinformatics-related information for teachers, being a complete source of information. Communication channels between researchers who use bioinformatics and teachers, are also provided in the webpage through a discussion forum for teachers to feel continuously supported.

Overall, we believe that "*Bioinformática na Sala de Aula*" may bring to the spotlight the educational benefits of bioinformatics as a didactic tool, capable to enhance the scientific and digital literacy of future active participants in society, i.e. students.

Acknowledgments

The authors are grateful to all the teachers who contributed with fruitful comments and feedback to the design and development of the webpage as well as to CIBIO-InBIO (https://cibio.up.pt/) and to Casa das Ciências (https://www.casadasciencias.org/) for the availability to share and promote this resource with the scientific and educational community. Ana Sofia Martins is supported by a fellowship from Fundação para a Ciência e Tecnologia – FCT (SFRH/BD/112038/2015).

References

Akpinar, Y., & Bayramoglu, Y. (2008). Promoting Teachers' Positive Attitude towards Web Use: A Study in Web Site Development. *Turkish Online Journal of Educational Technology, 7*(3). https://eric.ed.gov/?id=ED502675

Barker, L. (2009). Science teachers' use of online resources and the digital library for earth system education. *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries*, 1–10. https://doi.org/10.1145/1555400.1555402

Bikandi, J., Millán, R., Rementeria, A., & Garaizar, J. (2004). In silico analysis of complete bacterial genomes: PCR, AFLP–PCR and endonuclease restriction. *Bioinformatics, 20*(5), 798–799. https://doi.org/10.1093/bioinformatics/btg491

Bloom, M. (2001). Biology in silico: The Bioinformatics Revolution. *The American Biology Teacher, 63*(6), 397–403. https://doi.org/10.2307/4451145

Boulos, M., Maramba, I., & Wheeler, S. (2006). Wikis, blogs and podcasts: A new generation of Web-based tools for virtual collaborative clinical practice and education. *BMC Medical Education 6*(1), 41. https://doi.org/10.1186/1472-6920-6-41

Casa das Ciências. (2017). Bioinformática na Sala de Aula. Casa Das Ciências Divulga ... https://www.facebook.com/casa.das.ciencias.org/posts/1591259000934065

Caspi, R., Billington, R., Fulcher, C., Keseler, I., Kothari, A., Krummenacker, M., Latendresse, M., Midford, P., Ong, Q., Ong, W., Paley, S., Subhraveti, P., & Karp, P. (2017). The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Research, 46*(D1), D633–D639. https://doi.org/10.1093/nar/gkx935

CIBIO-InBIO. (2020). Extending bioinformatics to elementary and secondary school education. Teacher Training. https://cibio.up.pt/science-in-school-teacher-training/details/extending-bioinformatics-to-elementary-and-secondary-school-education

European Learning Laboratory for the Life Sciences. (2020). ELLS TeachingBase - Bioinformatics. ELLS European Learning Laboratory for the Life Sciences. http://emblog.embl.de/ells/teachingbase/?age=0&format=bioinformatics&lang=1

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. *PLoS Computational Biology 7*(10), e1002243. https://doi.org/10.1371/journal.pcbi.1002243

Hakverdi-Can, M., & Dana, T. (2012). Exemplary Science Teachers' Use of Technology. TOJET: *The Turkish Online Journal of Educational Technology, 11*(1).

Koszalka, T. (2001). Effect of Computer-Mediated Communications on Teachers' Attitudes Toward Using Web Resources in the Classroom. *Journal of Instructional Psychology, 28*(2), 95.

Kovarik, D., Patterson, D., Cohen, C., Sanders, E., Peterson, K., Porter, S., & Chowning, J. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education, 12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education, 26*(2), 195–206. https://doi.org/10.1080/0950069032000072782

Lewitter, F., & Bourne, P. E. (2011). Teaching Bioinformatics at the Secondary School Level. *PLoS Computational Biology, 7*(10), e1002242. https://doi.org/10.1371/journal.pcbi.1002242

Liaw, S.-S. (2002). Understanding user perceptions of World-wide web environments. *Journal of Computer Assisted Learning, 18*(2), 137–148. https://doi.org/10.1046/j.0266-4909.2001.00221.x

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible—the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics, 18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics, 14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Marques, I., Almeida, P., Alves, R., Dias, M., Godinho, A., & Pereira-Leal, J. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology, 10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M., Lemos, M., Lencastre, L., & Tavares, F. (2020). Bioinformatics-based activities in high school: Fostering students' literacy, interest and attitudes on gene regulation, genomics and evolution. *Frontiers in Microbiology, 11,* 578099. https://doi.org/10.3389/fmicb.2020.578099

Martins, A., Fonseca, M., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. The *American Biology Teacher, 80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2020). Bioinformatics, a Befitting Tool for e-learning: Potential and Constraints according Teachers' Perceptions. In M. F. Costa & J. B. Dorrío (Eds.), *Hands-on Science. Science Education*. Discovering and understanding the wonders of Nature (pp. 97–105). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2018). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. *3rd International Conference on Teacher Education* (INCTE), 203–214. http://hdl.handle.net/10198/17381

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In Manuel Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

Microbial Life. (2020). Bioinformatics in the Classroom: Online Resources. Microbial Life Educational Resources. https://serc.carleton.edu/microbelife/k12/bioinformatics/resources.html

Millán, R., Martínez-Ballesteros, I., Rementeria, A., Garaizar, J., & Bikandi, J. (2013). Online exercise for the design and simulation of PCR and PCR-RFLP experiments. *BMC Research Notes, 6*(1), 513. https://doi.org/10.1186/1756-0500-6-513

National Center for Biotechnology Information. (2020). Home - ORFfinder - NCBI. https://www.ncbi.nlm.nih.gov/orffinder/

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Network for Integrating Bioinformatics into Life Sciences Education. (2020). NIBLSE: Learning Resource Collection. Network for Integrating Bioinformatics into Life Sciences Education. https://qubeshub.org/community/groups/niblse/resourcecollection

Oner, D., & Adadan, E. (2011). Use of Web-Based Portfolios as Tools for Reflection in Preservice Teacher Education. *Journal of Teacher Education, 62*(5), 477–492. https://doi.org/10.1177/0022487111416123

Perrault, A. (2007). An Exploratory Study of Biology Teachers' Online Information Seeking Practices. American Library Association.

Richardson, W. (2010). Blogs, Wikis, Podcasts, and Other Powerful Web Tools for Classrooms. Corwin Press.

Romero, P., Wagg, J., Green, M. L., Kaiser, D., Krummenacker, M., & Karp, P. D. (2005). Computational prediction of human metabolic pathways from the complete human genome. *Genome Biology, 6*(1), R2. https://doi.org/10.1186/gb-2004-6-1-r2

Shaw, K., Horne, K., Zhang, H., & Boughman, J. (2008). Essay contest reveals misconceptions of high school students in genetics content. *Genetics*, *178*(3), 1157–1168. https://doi.org/10.1534/genetics.107.084194

Vallenet, D., Belda, E., Calteau, A., Cruveiller, S., Engelen, S., Lajus, A., Le Fèvre, F., Longin, C., Mornico, D., Roche, D., Rouy, Z., Salvignol, G., Scarpelli, C., Thil Smith,

A. A., Weiman, M., & Médigue, C. (2013). MicroScope--an integrated microbial resource for the curation and comparative analysis of genomic and metabolic data. *Nucleic Acids Research, 41*(D1), D636-47. https://doi.org/10.1093/nar/gks1194

Wheeler, S., & Wheeler, D. (2009). Using wikis to promote quality learning in teacher training. *Learning, Media and Technology, 34*(1), 1–10. https://doi.org/10.1080/17439880902759851

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School-New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology, 9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Yuen, S., Yaoyuneyong, G., & Yuen, P. (2011). Perceptions, interest, and use: Teachers and web 2.0 tools in education. *International Journal of Technology in Teaching and Learning, 7*(2), 109–123.

Zhang, M., Lin, C., Olsen, G., & Beck, B. (2007). A bioinformatics track with outreach components. *ITiCSE 2007: 12th Annual Conference on Innovation and Technology in Computer Science Education - Inclusive Education in Computer Science*, 186–190. https://doi.org/10.1145/1268784.1268840

Chapter VI

General Discussion

General Discussion

The new paradigms of biological research cannot be disregarded when seeking to promote a scientifically informed society. Indeed, they demand the need to improve curricular and educational resources at elementary and secondary education levels, through initiatives validated by focused science education research.

It has been reported that bioinformatics education action plans should be intended for students to get acquainted to these powerful resources and be encouraged to master a set of scientific tools that may be useful to address biology related questions (Dudley & Butte, 2009; Machluf & Yarden, 2013; Magana et al., 2014). Particularly in the last decade, we assisted to the release of several proposals to use bioinformatics in the classroom aiming to positively impact students learning, and enhance students' interest in scientific careers (STEM) (Amenkhienan & Smith, 2006; Kovarik et al., 2013; Machluf et al., 2017; Newman et al., 2016; Taylor et al., 2014). Despite these meritorious contributions, most of these proposals to bring bioinformatics to the classroom had to be tailored to the curricula requirements, the class schedules, validated by in-service teachers and, when possible, endorsed by their implementation in the classroom. Furthermore, these proposals missed the assessment needed to really evaluate the impact of bioinformatics-based activities on students' literacy, interest and motivation. This omission needs to be addressed in order to boost and strengthen bioinformatics integration at school in both formal and non-formal educational context (Magana et al., 2014).

It is undeniable that research institutions, wherein bioinformatics is part of the daily routine, are at a privileged position to propose a portfolio of introductory and user-friendly bioinformatics activities suitable for the elementary and secondary education curricular topics (Cummings & Temple, 2010; Ditty et al., 2010; Form & Lewitter, 2011; Wefer & Sheppard, 2008). Although, this cannot be accomplished without updating school curricula, nor by introducing core concepts, similarly to what has been done at the undergraduate level (Sayres et al., 2018; Welch et al., 2014).

As has been suggested by some studies, all educational stakeholders need to be recruited to scaffold bioinformatics dissemination in schools (Attwood et al., 2017; Koch & Fuellen, 2008). Nevertheless, teachers are acknowledged as holding a key role regarding the implementation of bioinformatics in school. In fact, teachers are pivotal educational agents to didactically explore bioinformatics to foster students' learning within a specific curricular context (Marques et al., 2014). Thus, it is urgent to raise the awareness of educational agents about the need to implement focused courses to

provide adequate training for teachers to confidently approach bioinformatics (Machluf et al., 2017; Marques et al., 2014).

Regardless these advancements, with the exception for sporadic initiatives (Machluf et al., 2017; Marques et al., 2014; Wood & Gebhardt, 2013), little has been done to prepare school teachers for this new paradigm. Elementary and secondary curricula and teaching practices are still organized according to formal models where institutional (schools) and individual (teachers) efforts are made to merge new information with previous knowledge aiming to foster students' cognition (Handelsman et al., 2004). The need to promote teacher training and empower teachers to adapt educational activities clashes with the overloaded curricula and the scarcity of bioinformatics-based didactic resources. To assertively intervene in teachers training it is mostly important to diagnose teachers' perceptions about bioinformatics integration in the school context, and comprehensively understand teachers needs to integrate bioinformatics in their schooling practices.

This PhD proposal aimed to be a robust contribution to upgrade elementary and secondary education concerning bioinformatics literacy, by acknowledging that teachers and students are core stakeholders, and taking into account the above-mentioned reality.

The data obtained during the course of this doctoral project contributed to better understand if bioinformatics tools and resources are influential to foster students' scientific and digital literacy, interest and attitudes and to depict which are the main constraints that are preventing teachers from fully engaging into the implementation of bioinformatics exercises in the classroom.

The research consisted in the conception, implementation and evaluation of a set of bioinformatics activities framed within elementary and secondary curricular contexts, using intuitive and open access bioinformatics resources to foster top-notch science education.

The outcomes of this doctoral project are detailed and discussed in the previous chapters of this thesis (Chapter II to Chapter V). The current chapter (Chapter VI), intends to offer an inclusive synopsis of the major findings and discuss its contributions to science education in general, and to the literacy on biology related themes, in particular, in view of the current knowledge on the topic, and to unfold its implications for the future research and practice.

1. "Ready…Set…Go!": *In silico* Tools Validated as Didactic Resources to Update Life Sciences Curricula in Middle and High School

When this PhD project started, a handful of bioinformatics-based activities, some of which combined with wet labs, were available as didactic instruments to taught contents of natural and life-science disciplines at the middle and high school level (Amenkhienan & Smith, 2006; Eurich et al., 2012; Gelbart & Yarden, 2006; Machluf et al., 2017; Marques et al., 2014; McQueen et al., 2012; Taylor et al., 2014). However, in these studies, three important variables detailed below were either neglected or not explicitly considered:

i) *time* – the proposals should be time-efficient, feasible within the class time-frame, and with explicit guidelines regarding how to drive the activity to ensure that all the exercises are carried out (Machluf & Yarden, 2013);

ii) *curriculum* – the activities should meet the ongoing curricula, to easy-up the integration of bioinformatics in the curricular topics, in order to be accepted by both teachers and students as an innovating teaching approach and a motivating learning environment, and not an extra burden (Form & Lewitter, 2011);

iii) *update concepts* - the integration of bioinformatics in the classroom should be an opportunity to reflect on the need to update the science curricula of middle and high school to the current societal demands and capable to address some diagnosed misconceptions, namely in what concerns genomics, genetics and evolution (Kovarik et al., 2013; Wefer & Sheppard, 2008).

The portfolio of 19 bioinformatics exercises proposed within the framework of this thesis, were conceived taking into account the above mentioned variables, namely the activities were designed attending the different school levels; the specific curricular topics (e.g. Molecular Biology, Gene Regulation, Mutations, Evolution, Bioinformatics in service to the population) (Appendix I); the time-frame planned of each class and curricular topic; and to the expected learning enhancement of core concepts (Martins & Tavares, 2018).

In this regard, in 2018 a proposal to update core concepts in high school genomics education was drawn and it is detailed in Chapter II (Martins & Tavares, 2018). It this study, it was shown that to fully understand the structure of genes, new core concepts, namely Open Reading Frame (ORF), and Basic Local Alignment Tool (BLAST) should be added to the current required notions of the Next Generation Science Standards

(National Research Council, 2013), or to the Portuguese national curricula (Mendes et al., 2003). Other concepts, such as synteny and comparative genomics, were shown to favor students' comprehension of bioinformatics activities and its potential to foster genomics education.

Among the exercises designed, a set of four bioinformatics-based exercises, supported by a tutorial video, to approach gene regulation, genomics and evolution is detailed in Chapter II (Martins, Fonseca, et al., 2018). These exercises, conceived to promote students' knowledge, motivation, critical thinking as well as their computational skills, fit the standards for elementary and high school curricula, which is essential for its successful implementation by biology teachers.

Summing up, four main deliverables were produced and are available for the educational community: a training course manual (Appendix I); workshop guidelines (Appendix II); and hands-on protocol, complemented with a tutorial video, aimed to identify genes and disclosing their genomic context (Martins, Fonseca, et al., 2018) and the relevance to introduce new genomics core concepts (Martins & Tavares, 2018).

2. "Ignition…Lift-off": Dry Labs Empowering Students' Scientific Toolbox

In order to integrate a new topic in the curricular practices, it is crucial to provide robust indicators on its positive impact on students' learning (both knowledge and skills). In the specific case of bioinformatics education, the lack on the assessment of its impact in elementary and secondary school level has been a major constraint to its integration in the classroom. Being aware that previous studies focused mainly on undergraduate and graduate level (Boyle, 2004; Brazas & Ouellette, 2016; Cattley & Arthur, 2007), and following recommendations that the curricular integration of bioinformatics should start earlier in the educational plan, to develop students' critical thinking and empower their decision-making competencies as future active citizens (Form & Lewitter, 2011; Gallagher et al., 2011; Lewitter & Bourne, 2011; Marques et al., 2014; McQueen et al., 2012; Porter & Smith, 2019; Wood & Gebhardt, 2013), this doctoral project intended to better characterize the learning impact of bioinformatics at elementary and secondary education. In line with this research, in the last few years more reports were released on the benefits of bioinformatics but mainly focused on formal learning (Boyles, 2019; Machluf et al., 2017; Minoli, 2018; Rueda et al., 2019; Whitley et al., 2020).

The assessment of the impact of bioinformatics activities on students' literacy, interest and attitudes in both formal and non-formal educational context is detailed in Chapter III of this thesis (Martins, Lencastre, et al., 2018a; Martins, Fonseca, et al., 2020). This

assessment caught attention for the importance to better align bioinformatics exercises with the curricula, to improve the guidelines of some bioinformatics-based exercises proposed and to validate further bioinformatics-based exercises to enrich the available portfolio.

Regarding the formal school settings, Martins, Fonseca, et al. (2020) developed research based on the impact of the activity "*Mining the genome: using bioinformatics tools in the classroom to support student discovery of genes*", on a sample of 387 students from 5 different schools, from different geographic regions of Portugal as described at Chapter III. This study allowed to observe that students are curious about bioinformatics and that after a theoretic introduction they feel more at ease to explore the bioinformatics platforms, getting acquainted with bioinformatics and acknowledge that numerous bioinformatics tools can be intuitive for beginners. Adding to this, activities contributed to improve students understanding about researchers' job and to perceive the importance of bioinformatics and computational biology to biological research, ultimately contributing to foster their motivation for STEM careers.

Furthermore, introducing comparative genomics in their learning practices contributed for a better understanding of curricular contents regarding identification of genes, their regulation and to make evolutionary assumptions. After the intervention students were able to pinpoint bioinformatics tools required to identify genes in genomics sequences and most importantly, they were able to resolve genomics-related misconceptions (Martins, Fonseca, et al., 2020). Overall, students revealed a positive attitude regarding the integration of bioinformatics-based approaches in their learning practices, reinforcing its added value in educational approaches and as a promoter of citizenship education. Taken together, these results highlight the potentialities of bioinformatics as an effective didactic tool to enhance students' skills and motivation, and most importantly to improve knowledge, namely by overcoming misconceptions and by the acquisition of new core concepts.

Complementing the assessment of bioinformatics impact in formal educational settings, a key contribution of this PhD project is related with non-formal educational context in which little was known regarding the learning added value of bioinformatics. During the summer of 2018 the activity named "*Bacteria, Antibiotics and Resistance: let´s find out the links*?" (Appendix III), included the use of a well-known food microbiology predictive platform designated ComBase to foster youths awareness about food safety and food preservation techniques (Martins, Lencastre, et al., 2018a). Following the assessment of this activity implemented under the framework of Porto Junior University as detailed in Chapter III, the authors observed that dry lab activities in

non-formal contexts have a positive impact on participants' knowledge, while promoting youngsters' citizenship education. In addition, it corroborates the results obtained in formal educational settings, particularly concerning pupils' positive perceptions about the importance of computer applications to biological research. This research stressed that bioinformatics framed within non-formal learning settings such as summer schools or science bootcamps contributes to positively impact students learning, while is particularly suitable to overcome some of the inherent limitations observed in formal learning setting, namely time and curricular constraints. The inexistence of strict curricular contents to comply with, makes these non-formal learning environments a playground to diversify proposals and combine *in silico* challenges with wet lab activities.

3. Bioinformatics as a Teaching Tool: An Extra Burden or a Trendsetter?

It is unquestionable to consider teachers as the main players of educational changes, and therefore the best research partners to develop effective tools adapted to schools' realities and students' needs (Machluf et al., 2012; Machluf & Yarden, 2013; Marques et al., 2014; Wefer & Sheppard, 2008; Wood & Gebhardt, 2013). In this regard, a major concern during this doctoral research project was to invite teachers to contribute to sculpt the pedagogical proposals based on bioinformatics tools. This effort to recruit teachers to the research was essential to characterize teachers' perceptions and attitudes towards bioinformatics education and identify the main constraints that are preventing teachers from fully engaging to implement bioinformatics exercises in the classroom. The data gathered concerning teachers' perspectives on bioinformatics integration was particularly instrumental to understand the feasibility and potential of bioinformatics education at different school levels in Portugal, where little has been done. These findings and all the data regarding teachers are detailed in Chapter IV (Martins et al., 2017; Martins, Lencastre, et al., 2018b, 2020c, 2020b).

To fully adequate teachers with competencies to implement bioinformatics-based activities in their teaching practices, a 25 hours training course for teachers designated "*Adequacy of Bioinformatics tools to Elementary Education and Secondary Education*" and a detailed training manual were launched in 2017 (Appendix I). At the end of this training, a diagnostic of teachers' perceptions about bioinformatics education and about acquired skills in basics bioinformatics was carried out as detailed in Chapter IV. Participant teachers were invited to fill in a questionnaire to determine several indicators that could contribute to better understand teachers' reality at schools - constraints and needs - as well as to validate the proposed exercises to afterwards implement with their

students. Teachers highlighted the curricular adequacy of the course contents, the user-friendly interfaces of the bioinformatics resources used and their perception of increased confidence to implement the bioinformatics activities in their classes. In contrast to these positive perceptions, teachers emphasized as the main constraints to implement bioinformatics at their classes the language barriers - English-based language of bioinformatics tutorials, also reported in previous studies (Machluf & Yarden, 2013), the multitude of commands characteristic of several bioinformatics applications and difficulties related with data analysis and interpretation (Martins et al., 2017). All the previous indicators obtained should be taken into consideration in other training actions.

To complement this study, it was also important to characterize the perceptions of teachers who did not attend the training course. In this regard, a group of 11 teachers which were willing to implement bioinformatics-based activities in their classes but that did not attend the training course were inquired. The results showed that these teachers followed the trend of the ones who attended the training course in which concerns the interest in bioinformatics, agreeing that it assumes a key role in biological research and highlighting the need of introducing curricular framed bioinformatics activities at elementary and secondary school levels (Martins, Lencastre, et al., 2018b) (Chapter IV). When teachers were asked about constraints to implement bioinformatics activities in the classroom, they mainly mentioned logistics problems, namely, poor internet access and lack of computers as the group of teachers who received training (Martins, Lencastre, et al., 2018b). Adding to this, all the teachers who did not attend the course showed interest in attending a training course in this scientific field. Interestingly, both groups of teachers agreed that "*The training courses available to address bioinformatics are still scarce*".

Lastly, we were also interested in acknowledging the impact of short-term (4h) training courses (workshops) on bioinformatics for teachers. In this scope, in 2018 a workshop for teachers entitled "*From DNA to Genes and to Comparative Genomics: Bioinformatics in the classroom*" was proposed within the program of a major annual meeting for science teachers (V Encontro Internacional da Casa das Ciências). As described above, follow up the workshop teachers were asked to fill an enquire to collect data regarding their previous training and academic background on bioinformatics, their attitudes towards bioinformatics integration and perceptions as described in Chapter IV (Martins, Lencastre, et al., 2020c). Teachers were shown to be motivated to attend short-term training sessions and interested in learning more about bioinformatics and about strategies to integrate bioinformatics in their classes, which makes sense considering that most of the teachers mentioned that their academic background was not sufficient

to confidently implement bioinformatics-based exercises in their classes. Teachers claimed for more and longer training courses in this area emphasizing the importance to foster more initiatives to integrate bioinformatics in secondary education curricula and highlighting the need to increase the offer of teachers' training on bioinformatics. Interestingly, these teachers admitted that their schools have the necessary conditions to implement bioinformatics-based approaches, which apparently contradicts the constraints mentioned by the group of teachers who attended the 25 hours training course, namely poor internet connection and lack of computers.

On 2019, a re-edition of the workshop was a key opportunity to clarify these findings and to increase the robustness of previous results. As detailed by Martins, Lencastre, et al. (2020b) in Chapter IV, generally, teachers acknowledged that their schools are equipped with computers and internet connection which may suggest that resources would be available to integrate bioinformatics in teaching practices. However, teachers admitted that often computers are obsolete with outdated software, poor internet connection and inaccessible for teaching. Focusing on these considerations the active use of educational web-based resources, in which bioinformatics can have an important role, calls for a digital reform of schools as encouraged by Next Generation Science Standards (NGSS) (National Research Council, 2013). It is also important to mention that teachers claimed for a longer training course (25h), instead of the workshop (4h) because they felt that it was too short to understand all the concepts and platforms disseminated. In fact, although the terms and the platforms used are not of a high level of complexity, they are new for teachers and in this regard, they need time to explore and to get acquainted to them in order to understand the full potential of these platforms as didactic tools. Altogether, these results underline the importance of teachers' training to enhance their skills and encourage innovation in their teaching practices.

The training course and workshops on bioinformatics for teachers lead to pinpoint teachers' competencies to take basic bioinformatics activities to their teaching practices.

Taking into consideration the genomics era we are living in it is increasingly acknowledged that bioinformatics education is unavoidable. Teachers need to be equipped with the competencies and tools to integrate this scientific area in their pedagogical practices, and mostly motivated to adequate basic bioinformatics resources to their teaching practices attending the curricula and aiming to increase students' literacy on genomics.

4. "*Keep in touch*": The Importance of the Closeness between Teachers and Research Institutions

Teachers have to be involved in the design of bioinformatics-based activities in order to validate them but also to raise a sense of ownership which is essential for teachers to successfully implement the activities in the classes (Machluf et al., 2012; Marques et al., 2014; Wefer & Sheppard, 2008; Whitley et al., 2020). Moreover, we acknowledged that educators claimed for more training courses and for continuous support provided by the academic and research institutions to continuously update and enhance their knowledge, interest and competencies on bioinformatics after the training course (Martins et al., 2017; Martins, Lencastre, et al., 2018b, 2020b; Wood & Gebhardt, 2013).

Some of the teachers clearly indicate that they need to feel supported, at least at an initial phase, in order to feel comfortable to take *in silico* tools to the classroom. In this regard, research centers and universities have a key role. There are institutions which already are aware of this need and have developed programs to give training to teachers and to support them directly at the school (Machluf & Yarden, 2013; Netherlands Bioinformatics Centre, 2009; Wood & Gebhardt, 2013). In Portugal, a similar initiative took place during 2014 but it was sporadic (Marques et al., 2014). More recently, within the framework of this PhD project, the webpage "*Bioinformática na Sala de Aula*" was created to scaffold teachers to integrate bioinformatics in their practices by meeting their reported needs, and to fuel networking between teachers, through a forum capable to stimulate the exchange of experiences regarding the use of bioinformatics applied to curricular contents (Martins, Lencastre, et al., 2020a) (Chapter V).

This webpage was designed taking into account biology teachers' self-reflections about the reasons preventing them to integrate bioinformatics in educational practices reveal, namely the need of training, the lack of time and the scarce offer of resources in Portuguese and focusing on scientific topics in which students' misconceptions were previously diagnosed. These issues were addressed by making available, guidelines of the activities for teachers and students; PowerPoint presentations for teachers; supplementary information; troubleshooting tips; and a discussion forum to foster teachers' networking. This webpage has been updated attending teachers' requests, and broadly divulged among teachers' community by partners such as Casa das Ciências. Despite the dissemination efforts made, unfortunately the access to the webpage has not been as frequent as desired (Chapter V), suggesting that new dissemination strategies have to be establish. In the future, it is expected to address the impact of the webpage in promoting bioinformatics-based learning.

Final Remarks

In 2016, when this PhD research project was initiated the knowledge regarding the impact of bioinformatics education in middle and high school was limited and mainly centered in formal educational settings. Adding to this, teachers' perspectives on the potential of this integration, namely in Portugal, needed to be better depicted. Moreover, hints on the core concepts related with genomics and bioinformatics that could contribute to overcome diagnosed misconceptions as well as to answer the call of updating life-science curriculum framed within the genomics era, were lacking. This PhD thesis gathers the findings obtained during the last four years framed within an updated knowledge and recent scientific developments, which I believe resulted in valuable advances:

- An inventory of updated genomics-related core concepts was established capable to instruct science curricula revisions (*Chapter II*).

- A portfolio of bioinformatics activities – curriculum-oriented, language-adapted (Portuguese), time-efficient, centered on overcoming misconceptions and updating genomics-related core concepts – was validated by in-service science teachers and implemented with students, being now available for science educators' community (*Chapter II*).

- The assessment of the potential of bioinformatics as a learning tool, namely its impact on students' knowledge, interest and attitudes, was performed in both formal and non-formal educational context. In this regard, data are available for both environments of teaching and learning. Adding to this, dedicated assessment tools (questionnaires) were designed, piloted, implemented and validated being now available for the educational community as a reliable and robust tool (*Chapter III*).

- A training course manual (25h) and a workshop handout (4h) were developed, validated and made available to teachers' community (Appendices I and II). Adding to these deliverables, the questionnaires created to collect data on teachers' academic background, sense of proficiency and attitudes towards bioinformatics integration are an instrument available to educational stakeholders to characterize teachers' perspective on the benefits of bioinformatics as a didactic resource, and may ultimately contribute to encourage biology curricula updates (*Chapter IV*).

- The diagnostic of teachers' perspectives, perceived-knowledge and self-reflections regarding bioinformatics gave a robust input to foster education in bioinformatics (*Chapter IV*).

- Students' and teachers' assessments provided valuable information to set a reference regarding the interventions needed to include bioinformatics in the Portuguese national frameworks (*Chapters III and IV*).

- The network with teachers by the designed long-term communications channels will certainly contribute for teachers to feel supported and it will make possible to follow-up the experiences when implementing the proposed bioinformatics resources (Chapter V).

Future Perspectives

I believe that the findings and deliverables of this study may now serve as the foundation for further studies in this topic, capable to address several questions raised within the scope of this PhD project, namely:

- **Strengthen the Portfolio of Bioinformatics-Based Activities**: During this doctoral project, as described before, a portfolio of bioinformatics labs was organized as didactics resources both for middle and high school level, in formal or non-formal learning settings. However, only some of these activities were implemented in a classroom context. Therefore, it will be important to implement and validate all the bioinformatics-based activities proposed, particularly focusing on exploring other curricular topics and tackling other misconceptions that were not objects of intervention within the current project. Furthermore, we recently acknowledged the interest of schoolbooks editors to publish the portfolio of bioinformatics-based activities developed within the framework of this PhD project, at the e-learning platforms of the editor, which will increase the visibility of the project.

- **Edition of Didactic Tutorial Videos**: The implementation of bioinformatics-based activities carried out in school settings during the 2nd year of the project, suggested that the edition of tutorial videos is particularly welcome in comparison with detailed written guidelines to which students and teachers are generally less receptive. New tutorial videos will certainly contribute to successfully execute more activities.

- **Extending and Diversifying the Students' Populations Studied**: For future investigations, it will be important to increase the sample universe to allow robust comparisons with the data obtained in this study. In addition, it will be important to extend these studies across several years to understand the long-term effect of the interventions carried out.

- **The Future Challenges of Teachers' Training:** Reediting the training course on bioinformatics for teachers and promoting new workshops will lead to draw solid conclusions, regarding teachers' competencies to take basic bioinformatics activities to their teaching practices. This is particularly relevant taking into account the null or poor

background of teachers in bioinformatics which results in a lack of confidence to implement basic bioinformatics exercises in the classroom. Another important dimension to explore in the near future is the integration of introductory bioinformatics in the curricular contents of pre-service science teachers' courses and understand the receptivity and the potentialities of *in silico* tools seen by the eyes of the future science teachers. During this PhD research, it was particularly evident the high motivation of in-service teachers to integrate bioinformatics in their practices in order to take the "top-notch" research to their students, although the motivation revealed by pre-service teachers remains to be elucidate. This is a field which requires further study having in mind the need to update teachers practices in the future which seems not to be followed by pre-service teachers.

- **Teacher Influence in Students' Performance**: In a future study, a nested effect could help to depict the influence of the "*teacher*" in students' performance. Complementing this analysis, a cross-comparison between students' and teachers' perceptions about bioinformatics and bioinformatics education would supplement the results obtained in this research. For example, characterizing students' perceptions of learning and the effective learning after attending a class with a trained-on bioinformatics teacher (25h) and after attending a class with a teacher which got a short time training on bioinformatics (4h). Also, once students revealed to be aware that in-service teachers could not be acquainted with computers and technology in the class, it would be important to cross-examine students' perceptions with teachers perceived skills on technological tasks.

- **Webpage Dissemination Strategy**: Although the webpage "*Bioinformática na Sala de Aula*" was launched in 2017, no formal assessment of the use of this webpage by teachers was performed yet. Nevertheless, attending the limited accesses to the webpage, mainly carried out by teachers who attend bioinformatics training or who collaborated in the research, it is important to better disseminate the webpage using more proficient strategies. Designing an e-learning training course in which teachers have to access the webpage to accomplish specific tasks (as download the handouts, "wikis", podcasts or to use the forum) could be a powerful strategy. The webpage improvements, resulting from teachers' feedback, which takes time, will certainly encourage more teachers to use bioinformatics resources in their practices. The impact that the webpage may have in promoting bioinformatics-based learning will allow gather further information to depict its influence and to improve the webpage aiming to meet educators' needs.

- **The Importance of Digital-based Learning after December 2019**: Lastly, it is inescapable to make brief considerations on the meaning that the research developed within the framework of this PhD project may have regarding the recent pressure to rapidly implement *e*-learning strategies to attend the compulsory containment due to COVID-19 pandemic. In fact, bioinformatics as an inter- and transdisciplinary didactic tool, with trustworthy benefits for students' scientific education, is at a privileged position to meet the recent recommendation made by the Portuguese Ministry of Science, technology and Higher Education for the preparation of the academic year 2020/2021 (Gabinete do Ministro da Ciência Tecnologia e Ensino Superior, 2020). In this document it is particularly emphasized "*to stimulate innovation and pedagogical modernization (…) innovative teaching and learning adapted to a classroom-based education system supported digital technologies, as well as mixed / combined forms of education in all higher education broadening and deepening project-based ways of learning and teaching, integration of self-learning and teamwork* (…)". The same word and stream of consciousness are fully adequate to middle and secondary school level in order to meet the challenges that educational society is facing and in which bioinformatics-based education can have a key role.

References

Amenkhienan, E., & Smith, E. (2006). A web-based genetic polymorphism learning approach for high school students and science teachers. *Biochemistry and Molecular Biology Education*, *34*(1), 30–33. https://doi.org/10.1002/bmb.2006.49403401030

Attwood, T., Blackford, S., Brazas, M., Davies, A., & Schneider, M. (2017). A global perspective on evolving bioinformatics and data science training needs. *Briefings in Bioinformatics, 20*(2), 398-404. https://doi.org/10.1093/bib/bbx100

Boyle, J. (2004). Bioinformatics in undergraduate education: Practical examples. *Biochemistry and Molecular Biology Education*, *32*(4), 236–238. https://doi.org/10.1002/bmb.2004.494032040376

Boyles, F. (2019). *Bringing practical bioinformatics to high school classrooms*. Oxford Protein Informatics Group. https://www.blopig.com/blog/2019/09/bringing-practical-bioinformatics-to-high-school-classrooms/

Brazas, M., & Ouellette, B. (2016). Continuing Education Workshops in Bioinformatics Positively Impact Research and Careers. *PLoS Computational Biology*, *12*(6), e1004916. https://doi.org/10.1371/journal.pcbi.1004916

Cattley, S., & Arthur, J. (2007). BioManager: the use of a bioinformatics web application as a teaching tool in undergraduate bioinformatics training. *Briefings in*

Bioinformatics, *8*(6), 457–465. https://doi.org/10.1093/bib/bbm039

Cummings, M., & Temple, G. (2010). Broader incorporation of bioinformatics in education: Opportunities and challenges. *Briefings in Bioinformatics*, *11*(6), 537–543. https://doi.org/10.1093/bib/bbq058

Ditty, J., Kvaal, C., Goodner, B., Freyermuth, S., Bailey, C., Britton, R., Gordon, S. G., Heinhorst, S., Reed, K., Xu, Z., Sanders-Lorenz, E., Axen, S., Kim, E., Johns, M., Scott, K., & Kerfeld, C. (2010). Incorporating Genomics and Bioinformatics across the Life Sciences Curriculum. *PLoS Biology*, *8*(8), e1000448. https://doi.org/10.1371/journal.pbio.1000448

Dudley, J., & Butte, A. (2009). A Quick Guide for Developing Effective Bioinformatics Programming Skills. *PLoS Computational Biology*, *5*(12), e1000589. https://doi.org/10.1371/journal.pcbi.1000589

Eurich, C., Fields, P., & Rice, E. (2012). Proteomics: Protein Identification Using Online Databases. *The American Biology Teacher*, *74*(4), 250–255. https://doi.org/10.1525/abt.2012.74.4.8

Form, D., & Lewitter, F. (2011). Ten simple rules for teaching bioinformatics at the high school level. *PLoS Computational Biology*, *7*(10), e1002243. https://doi.org/10.1371/journal.pcbi.1002243

Gabinete do Ministro da Ciência Tecnologia e Ensino Superior. (2020). *Recomendação às Instituições Científicas e de Ensino Superior para a preparação do Ano Letivo 2020/2021*.

Gallagher, S., Coon, W., Donley, K., Scott, A., & Goldberg, D. (2011). A first attempt to bring computational biology into advanced high school biology classrooms. *PLoS Computational Biology*, *7*(10), e1002244. https://doi.org/10.1371/journal.pcbi.1002244

Gelbart, H., & Yarden, A. (2006). Learning genetics through an authentic research simulation in bioinformatics. *Journal of Biological Education*, *40*(3), 107–112. https://doi.org/10.1080/00219266.2006.9656026

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R., Gentile, J., Lauffer, S., Stewart, J., Tilghman, S. M., & Wood, W. B. (2004). Scientific Teaching. *Science*, *304*(5670), 521–522. https://doi.org/10.1126/science.1096022

Koch, I., & Fuellen, G. (2008). A review of bioinformatics education in Germany. *Briefings in Bioinformatics*, *9*(3), 232–242. https://doi.org/10.1093/bib/bbn006

Kovarik, D., Patterson, D., Cohen, C., Sanders, E., Peterson, K., Porter, S., &

Chowning, J. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, *12*(3), 441–459. https://doi.org/10.1187/cbe.12-11-0193

Lewitter, F., & Bourne, P. (2011). Teaching Bioinformatics at the Secondary School Level. *PLoS Computational Biology*, *7*(10), e1002242. https://doi.org/10.1371/journal.pcbi.1002242

Machluf, Y., Gelbart, H., & Yarden, A. (2012). High-School Teachers' Appropriation of an Innovative Curriculum in Bioinformatics. In D. Kruger & M. Ekborg (Eds.), *The 9th Conference of European Researchers in Didactics of Biology (ERIDOB)*.

Machluf, Y., & Yarden, A. (2013). Integrating bioinformatics into senior high school: design principles and implications. *Briefings in Bioinformatics*, *14*(5), 648–660. https://doi.org/10.1093/bib/bbt030

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2017). Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Briefings in Bioinformatics*, *18*(1), 145–159. https://doi.org/10.1093/bib/bbv113

Magana, A., Taleyarkhan, M., Alvarado, D., Kane, M., Springer, J., & Clase, K. (2014). A Survey of Scholarly Literature Describing the Field of Bioinformatics Education and Bioinformatics Educational Research. *CBE—Life Sciences Education*, *13*(4), 607–623. https://doi.org/10.1187/cbe.13-10-0193

Marques, I., Almeida, P., Alves, R., Dias, M. J., Godinho, A., & Pereira-Leal, J. (2014). Bioinformatics Projects Supporting Life-Sciences Learning in High Schools. *PLoS Computational Biology*, *10*(1), e1003404. https://doi.org/10.1371/journal.pcbi.1003404

Martins, A., Fonseca, M. J., Lemos, M., Lencastre, L., & Tavares, F. (2020). Bioinformatics-based activities in high school: Fostering students' literacy, interest and attitudes on gene regulation, genomics and evolution. *Frontiers in Microbiology, 11,* 578099. https://doi.org/10.3389/fmicb.2020.578099

Martins, A., Fonseca, M. J., & Tavares, F. (2018). Mining the Genome: Using Bioinformatics Tools in the Classroom to Support Student Discovery of Genes. *The American Biology Teacher*, *80*(8), 619–624. https://doi.org/10.1525/abt.2018.80.8.619

Martins, A., Lencastre, L., & Tavares, F. (2017). Adequacy of bioinformatics tools to elementary and secondary school curricula: a training course for teachers. In *Congresso Internacional - O Tempo dos Professores*. FPCEUP. https://sigarra.up.pt/fpceup/pt/pub_geral.pub_view?pi_pub_base_id=262743&pi_pub_r

1_id=

Martins, A., Lencastre, L., & Tavares, F. (2018a). Predictive Microbiology in a Non-Formal Science Education Context: Understanding Food Preservation Techniques. In M Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 309–317). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2018b). Integrating bioinformatics in elementary and secondary education: Teacher's perceptions. In *3rd International Conference on Teacher Education (INCTE)*. Instituto Politécnico de Bragança. http://hdl.handle.net/10198/17381

Martins, A., Lencastre, L., & Tavares, F. (2020a). "Bioinformática na Sala de Aula": Webpage to Boost Bioinformatics in the Classroom. *Simpósio Internacional de Psicologia Da Educação: Passado, Presente e Futuro (SInPE20)*.

Martins, A., Lencastre, L., & Tavares, F. (2020b). Bioinformatics, a Befitting Tool for e-learning: Potential and Constraints according Teachers' Perceptions. In M. F. Costa & J. B. Dorrío (Eds.), *Hands-on Science. Science Education. Discovering and understanding the wonders of Nature* (pp. 97–105). Hands-on Science Network.

Martins, A., Lencastre, L., & Tavares, F. (2020c). Bringing Bioinformatics to Secondary Education: A Workshop for Science Teachers. *The Beauty and Pleasure of Understanding: Engaging with Contemporary Challenges Through Science Education*. European Science Education Research Association. *In Press.*

Martins, A., & Tavares, F. (2018). Genomics Education: Update Core Concepts in High School. In Manuel Costa, B. Dorrío, & J. Fernandez-Novell (Eds.), *Hands-on Science. Advancing Science. Improving Education.* (pp. 145–150). Hands-on Science Network.

McQueen, J., Wright, J., & Fox, J. (2012). Design and Implementation of a Genomics Field Trip Program Aimed at Secondary School Students. *PLoS Computational Biology*, *8*(8), e1002636. https://doi.org/10.1371/journal.pcbi.1002636

Mendes, A., Rebelo, D., & Pinheiro, E. (2003). *Programa de Biologia e Geologia 11º ou 12ºano(s)*. Ministério da Educação: Departamento do Ensino Secundário.

Minoli, M. (2018, January 9). *Didactic research in High School for innovative STEM bioinformatics activities « Scientix blog*. Scientix - The Community for Science Education in Europe. https://blog.scientix.eu/2018/01/didactic-research-in-high-school-for-innovative-stem-bioinformatics-activities/

National Research Council. (2013). *Next Generation Science Standards*. National

Academies Press. https://doi.org/10.17226/18290

Netherlands Bioinformatics Centre. (2009). *Bioinformatics as a didactic tool in high school*. NBIC Teacher Training. https://www.nbic.nl/education/high-school-programmes/bioinformaticsschool/teacher-training/index.html

Newman, L., Duffus, A., & Lee, C. (2016). Using the Free Program MEGA to Build Phylogenetic Trees from Molecular Data. *The American Biology Teacher*, *78*(7), 608–612. https://doi.org/10.1525/abt.2016.78.7.608

Porter, S., & Smith, T. (2019). Bioinformatics for the Masses: The Need for Practical Data Science in Undergraduate Biology. *Omics : A Journal of Integrative Biology*, *23*(6), 297–299. https://doi.org/10.1089/omi.2019.0080

Sayres, M., Hauser, C., Sierk, M., Robic, S., Rosenwald, A. G., et al. (2018). Bioinformatics core competencies for undergraduate life sciences education. *PLoS ONE*, *13*(6), e0196878. https://doi.org/10.1371/journal.pone.0196878

Rueda, A., Benítez, G., Marchetti, J., Hasenahuer, M., Fornasari, M., Palopoli, N., & Parisi, G. (2019). Bioinformatics calls the school: Use of smartphones to introduce Python for bioinformatics in high schools. *PLoS Computational Biology*, *15*(2), e1006473. https://doi.org/10.1371/journal.pcbi.1006473

Taylor, J., Davidson, R., & Strong, M. (2014). Drug-resistant tuberculosis: A genetic analysis using online bioinformatics tools. *American Biology Teacher*, *76*(6), 386–394. https://doi.org/10.1525/abt.2014.76.6.6

Wefer, S., & Sheppard, K. (2008). Bioinformatics in high school biology curricula: A study of state science standards. *CBE Life Sciences Education, 7*(1), 155-162. https://doi.org/10.1187/cbe.07-05-0026

Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., & Schneider, M. (2014). Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. *PLoS Computational Biology*, *10*(3), e1003496. https://doi.org/10.1371/journal.pcbi.1003496

Whitley, K., Tueller, J., & Weber, K. (2020). Genomics Education in the Era of Personal Genomics: Academic, Professional, and Public Considerations. *International Journal of Molecular Sciences*, *21*(3), 768. https://doi.org/10.3390/ijms21030768

Wood, L., & Gebhardt, P. (2013). Bioinformatics Goes to School—New Avenues for Teaching Contemporary Biology. *PLoS Computational Biology*, *9*(6), e1003089. https://doi.org/10.1371/journal.pcbi.1003089

Appendices

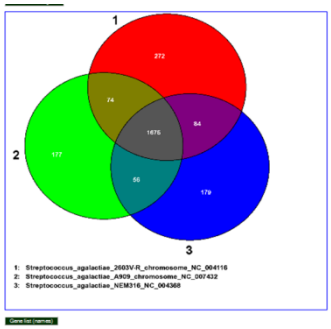
Appendix I: Training Couse Manual

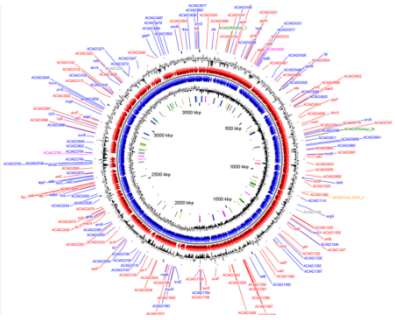
Adequação de Ferramentas Bioinformáticas ao 3ºCiclo do Ensino Básico e ao Ensino Secundário (CCPFC/ACC-88413/16): Training course manual

Curso de Formação (CCPFC/ACC-88413/16)

Adequação de Ferramentas Bioinformáticas ao 3ºCiclo do Ensino Básico e ao Ensino Secundário

Faculdade de Ciências, Universidade do Porto





Fernando Tavares | Pedro Albuquerque | Cláudia Serra

Ana Chung | Maria João Fonseca | Ana Martins

Listagem de Abreviaturas e Siglas

BLAST - Basic Local Alignment Search Tool: esta ferramenta permite, através de um algoritmo matemático, localizar as sequências mais semelhantes à sequência em estudo que se encontram presentes na base de dados.

FASTA - formato baseado em texto para representar tanto sequências de nucleótidos como sequências de peptídeos, no qual os nucleotídeos ou aminoácidos são representados usando códigos de uma única letra.

MaGe - Microbial Genome Annotation & Analysis Platform: https://www.genoscope.cns.fr/agc/microscope/home/index.php

NCBI - National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/

ORF - Open reading frame: sequência de DNA compreendida entre um codão de início (geralmente ATG) da tradução e um codão de terminação, descontando as sequências que correspondem aos intrões no caso de os haver.

PCR - Polymerase Chain Reaction: Amplificação de uma sequência de DNA por meio de ciclos térmicos de desnaturação da cadeia de DNA alvo, seguida da replicação usando oligonucleotídeos de iniciação (*primers*) e uma DNA polimerase termoestável (geralmente *Taq* polimerase).

SynMap- Permite gerar mapas de sintenia entre dois organismos e identificar regiões homólogas.

TCA cycle - tricarboxylic acid (TCA) cycle: Ciclo de Krebs ou Ciclo do Ácido tricarboxilico.

Introdução

Os desafios e mudanças a que atualmente assistimos no ensino das Ciências exige dos professores uma adaptação e uma necessidade de exploração de abordagens inovadoras que permitam comprometer os alunos com um trabalho científico alicerçado na resolução de problemas e na formulação de hipóteses. A tarefa dos professores é dar a conhecer a Ciência segundo uma perspetiva realista e atual, atendendo sempre à motivação como um fator essencial no processo de aprendizagem (Council, 2003).

Neste sentido, surge a oferta deste curso de formação que terá por base a bioinformática como ferramenta de ensino, sendo o principal objetivo da ação o contacto com ferramentas bioinformáticas específicas, enquadradas curricularmente e que constituem recursos digitais com enorme potencial.

A bioinformática como disciplina científica surgiu em resposta à necessidade de sistematizar de forma eficaz e expedita a enorme quantidade de dados obtidos experimentalmente, tendo por base a utilização de ferramentas computacionais para adquirir, analisar e armazenar, informação biológica contida essencialmente nos ácidos nucleicos e proteínas, de forma a elaborar hipóteses e alcançar interpretações de mecanismos moleculares e celulares que levem a um melhor conhecimento da biologia dos organismos (Madigan M., 2010).

Para além de permitir uma abordagem na área das Ciências Naturais e da Biologia, a bioinformática é também uma disciplina que intersecta outras áreas relacionadas como a ciência de computadores, a matemática e até mesmo a física e a química, sendo por isso interdisciplinar e integradora, facilitando agendas de colaboração entre professores de áreas disciplinares distintas, e permitindo mobilizar alunos com apetências e motivações complementares. De facto, a bioinformática assume um papel relevante no aumento da literacia digital dos alunos, permitindo alargar a sua "caixa de ferramentas científicas", e indo ao encontro das competências definidas nas metas curriculares para as Tecnologias da Informação e da Comunicação de que são exemplos: o uso das tecnologias para difundir a literacia digital, a análise crítica da informação disponível através das tecnologias e a estimulação dos alunos como utilizadores ativos dos computadores, da Internet e das redes de colaboração (Dudley & Butte, 2009; Horta M., 2012). A literacia digital pode ser descrita como a capacidade de encontrar, avaliar, utilizar, compartilhar e criar conteúdos usando tecnologias de informação e da Internet. Esta é uma vertente fundamental para a Sociedade do Conhecimento na qual os alunos são agentes ativos, sendo esta iniciativa um passo em frente na incorporação desta temática na educação (Ilomäki, Paavola, Lakkala, & Kantosalo, 2016; University, 2009; Zylka, Christoph, Kroehne, Hartig, & Goldhammer, 2015). Adicionalmente, atividades que proporcionam o contacto com recursos usados em laboratórios de investigação sabe-se serem um estímulo ao interesse dos alunos pelas carreiras científicas (STEM), uma vez que permitem aos discentes conhecer as principais funções desta área profissional e saber quais os requisitos de formação necessários (Kovarik, 2013).

Algumas iniciativas ocorridas noutros países têm sido desenvolvidas com vista a tentar integrar a bioinformática nos currículos de outros países, de que são exemplo atividades centradas no ensino da evolução ou da expressão genética. Urge então a necessidade de atualização das práticas docentes nas nossas salas de aula sem entrar em conflito com os planos curriculares (Amenkhienan & Smith, 2006; Boyle, 2004; Caroline Alexandra, 2001).

Atendendo ao contexto internacional, estudos recentes apontam para um crescente envolvimento dos professores na aplicação da bioinformática nas suas salas de aula. No entanto, é também descrita a necessidade fundamental da frequência de cursos de formação nesta área, para que os professores possam explorar com confiança temas e ferramentas com um considerável grau de exigência. A corroborar esta prioridade existem dados recentes que revelam que, após uma primeira incursão em atividades de bioinformática, os professores

tendem a sentir-se cada vez mais motivados, confiantes e capazes de redesenhar os objetivos didático pedagógicos a alcançar (Machluf, Gelbart, Ben-Dor, & Yarden, 2016).

Alicerçada na aplicação de ferramentas bioinformáticas como recurso de trabalho para os professores, o curso de formação *"Adequação de ferramentas Bioinformáticas ao 3ºciclo do Ensino Básico e ao Ensino Secundário"* visa ainda uma componente de contextualização curricular fundamental que resultou da análise das orientações oficiais referentes às Ciências Naturais do 3º ciclo do Ensino Básico e de Biologia do Ensino Secundário, com o objetivo de apresentar soluções que vão de encontro às exigências dos programas curriculares vigentes. As temáticas abordadas nesta ação de formação estão enquadradas em vários níveis de escolaridade, nomeadamente nos programas das disciplinas de Ciências Naturais do 8ºano e 9ºano, de Biologia e Geologia do 10º e 11º anos e de Biologia do 12º ano, de acordo com a planificação descrita na secção "Conteúdos da Ação".

A frequência desta ação de formação permitirá aos formandos consolidar conhecimentos, adquirir conceitos emergentes que surgiram com o desenvolvimento da Bioinformática, e desenvolver novas competências. No seu conjunto, esta ação de formação deverá habilitar os formandos para a implementação de um ensino baseado em atividades dry lab, com base na exploração de recursos *in silico*.

Deve ainda ser salientado, que os protocolos desenhados especificamente para esta ação são baseados em aplicações e bases de dados bioinformáticas de acesso livre, possibilitando a sua utilização e exploração continuada pelos formandos, sem encargos materiais e permitindo uma calendarização que seja compatível com as disponibilidades letivas. Esta segunda edição do Guia do Curso de Formação beneficia de melhoramentos, decorrentes da 1ªedição do Curso de Formação (março-abril 2017), no que se refere à elaboração e estrutura dos exercícios e à inclusão de novas atividades.

Objetivos do Curso

Esta ação de formação tem com objetivo geral responder à necessidade de atualização dos conhecimentos e práticas dos professores de Ciências Naturais do 3ºciclo do Ensino Básico e de Biologia do Ensino Secundário. Recorrendo a ferramentas de bioinformática, de livre acesso e com interfaces intuitivas usadas diariamente em investigação, os professores serão deparados com um cenário de ensino que fomentará e divulgará práticas pedagógicas inovadoras e motivadoras para aplicar nas suas salas de aula, sem custos associados.

No âmbito deste objetivo geral, são ainda objetivos desta formação:

- Desenvolver a literacia digital e fornecer recursos para a sua estimulação e enriquecimento em sala de aula;
- Contribuir para o desenvolvimento de atitudes positivas face às novas tecnologias e à sua utilização como ferramenta de ensino das Ciências;
- Sensibilizar para a importância do trabalho prático como parte integrante e fundamental do processo de ensino-aprendizagem;
- Fomentar a aquisição de conhecimentos, instrumentos e metodologias necessários à aplicação e ao desenvolvimento de novas atividades práticas;
- Proporcionar a utilização de materiais e ferramentas bioinformáticas que possam ser particularmente úteis em contexto escolar;
- Promover o recurso a metodologias de ensino ativas, com carácter *inquiry-based*, que têm como principal objetivo envolver os alunos no processo de investigação científica, contribuindo para o aumento da sua motivação perante as tarefas a realizar e fomentando o desenvolvimento de capacidades de raciocínio e de pensamento crítico e criativo;
- Incentivar os formandos a estabelecer parcerias entre as escolas do ensino básico e secundário e as instituições de ensino superior/investigação, no sentido de melhorar as ações de cooperação entre as mesmas e facilitar o desenvolvimento e a transferência de práticas inovadoras para a melhoria e qualidade do ensino.

Conteúdo Programático

A ação de formação está organizada para um número total de 25 horas presenciais, distribuídas em sete sessões teórico-práticas: cinco sessões de 4 horas e duas sessões com a duração de 2,5 horas.

Sessão 1 (2,5h)
- Abertura e apresentação do curso de formação: objectivos; cronograma e critérios de avaliação.
- Introdução à Bioinformática: Apresentação, objetivos e descrição sumária de várias bases de dados e aplicações bioinformáticas. Exemplos da utilização da bioinformática em investigação.

Sessão 2 (4h) - "Biologia Molecular: Análises *in silico*"

- Parte I: A primeira parte da sessão visa a análise das enzimas de restrição e o seu funcionamento, assim como a simulação *in silico* de experiências com recurso à técnica de PCR, recorrendo à plataforma bioinformática *In silico simulation of molecular biology experiments* (http://insilico.ehu.es/). Apesar das potencialidades desta plataforma serem mais alargadas, será dado ênfase a duas aplicações que vão de encontro a tópicos curriculares atuais: i) Amplificação por PCR, tendo como objetivo a reconstituição e confirmação *in silico* de uma experiência de diagnóstico de bactérias patogénicas reportada num artigo científico; e ii) A elaboração de mapas de restrição de genomas de diferentes estirpes bacterianas com diferentes enzimas de restrição. As componentes desta abordagem vão ao encontro das orientações curriculares do programa de Biologia do 12ºano, essencialmente materializadas no tema 4 - Alterações do Material Genético: 4.2. Fundamentos da Engenharia Genética, em que são abordadas matérias como a importância das enzimas de restrição como ferramentas da engenharia genética.

- Parte II: Numa segunda parte da sessão serão exploradas ferramentas de análise de sequências de DNA. Recursos bioinformáticos, nomeadamente o *NCBI ORF finder* (http://www.ncbi.nlm.nih.gov/orffinder/) e o chamado BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), serão usados para análise *in silico* de um fragmento de DNA cuja função é desconhecida. Será possível identificar os genes putativos e suas funções, comparando com sequências já conhecidas. Este exercício permitirá explorar ao nível do 12ºano de escolaridade a organização do material genético, enfatizando, por exemplo, os cromossomas como entidades que contêm genes ou o papel dos operões para os seres procariontes e dos componentes intervenientes em mecanismos de regulação (Módulo 1: Reprodução e Património Genético – 3.2. Organização e regulação do material genético). Este conhecimento contribuirá para uma abordagem holística de noções básicas como genoma, gene e a importância de regiões intergénicas.

Sessão 3 (4h) – "Operão *lac*: Regulação Génica e Relações Evolutivas"

- Conclusão de tarefas pendentes da sessão 2.

- Parte I: É objetivo da primeira parte desta sessão a identificação de codões de iniciação. Analisando o código genético é possível encontrar diferentes codões de iniciação. Com auxílio da aplicação de análise de ORF's NCBI ORF *finder* (http://www.ncbi.nlm.nih.gov/orffinder/) será possível abordar temas enquadrados no 12ºano de escolaridade como a expressão genética, podendo também ser dinamizados no 11ºano de escolaridade com o objetivo de explorar os processos de transcrição e de tradução (Unidade 5: Crescimento e renovação celular – 1.1. DNA e Síntese Proteica), interpretando qual o possível ou possíveis codões de iniciação de um determinado gene (*lac I*).

- Parte II: Na segunda parte da sessão será utilizada uma plataforma bioinformática de genómica comparativa (MaGe – *MicroScope* - https://www.genoscope.cns.fr/agc/microscope/home/index.php) sendo possível investigar relações evolutivas, identificando a presença dos genes do operão *lac* e das regiões flanqueantes em diferentes grupos taxonómicos. Este exercício permitirá reforçar a consistência das conclusões obtidas na sessão 2 quanto à função dos genes, mas também elaborar hipóteses evolutivas que possam explicar a presença dos genes em taxa distintos.

Sessão 4 (4h) – "Importância dos Genes e das Mutações no Estudo da Evolução"

- Parte I: No início da quarta sessão propõe-se a utilização de ferramentas bioinformáticas e bases de dados disponibilizados no EDGAR (https://edgar.computational.bio.uni-giessen.de/). Escolhendo até cinco estirpes bacterianas representantes dos grupos taxonómicos analisados na Parte II da sessão 3, pretende-se identificar o conjunto de genes homólogos e genes específicos de cada estirpe utilizando a funcionalidade "Diagramas de Venn" na plataforma EDGAR. Com base nos resultados obtidos discutir as noções de genoma centro (*core* genoma), *pan* genoma e genoma acessório. O EDGAR permitirá ainda identificar a sequência nucleotídica de cada gene e inferir a sequência aminoacídica, tendo em consideração a frequência de utilização de codões para diferentes organismos (consultando codon usage table).Esta proposta enquadra-se particularmente no 11ºano de escolaridade na Unidade 5: Crescimento e renovação celular– 1.1. DNA e Síntese Proteica; Unidade 8: Sistemática dos Seres Vivos - 1.2. Taxonomia e Nomenclatura.

- Parte II: É objetivo da segunda parte desta sessão estabelecer *plots* e mapas de sintenia através da comparação de genomas de duas das estirpes analisadas na Parte I desta mesma sessão, utilizando essencialmente o SynMap da plataforma CoGe (https://genomevolution.org/CoGe/). Neste exercício para além da noção de sintenia, os formandos poderão observar e registar importantes eventos de dinâmica genómica, nomeadamente inversões, deleções e inserções. Permite-se assim a abordagem a conceitos de evolução enquadrados no 11ºano de escolaridade (Unidade 7 – 2. Mecanismos de Evolução), mas também relativos à compreensão da utilidade das mutações na exploração do tema 4. Alteração do material genético – 4.1. Mutações (Módulo 1) do programa 12ºano de escolaridade.

Sessão 5 (4h) – "Explorando Vias Metabólicas nos Diferentes Domínios da Vida"

- Parte I: É objetivo da parte I desta sessão a exploração da ferramenta de análise MetaCyc - MetaCyc Metabolic Pathway Database (http://metacyc.org/) que permitirá o estudo de reações do ciclo de Krebs (TCA cycle), e dos compostos envolvidos, com vista a identificar diferenças neste ciclo para diferentes domínios da vida, como os Eukarya e Bacteria. Este exercício será ainda um valioso contributo para compreender o carácter anfibólico dos compostos intermediários do ciclo de Krebs e o seu envolvimento noutras reações do metabolismo celular. Com este exercício os formandos poderão fundamentar a biossíntese de moléculas precursoras de outras biomoléculas essenciais como aminoácidos, purinas, pirimidinas e ácidos gordos. Esta atividade enquadra-se curricularmente no 10ºano de escolaridade: Unidade 3 – Transformação e utilização de energia pelos seres vivos: 1. Fermentação; 2. Respiração aeróbia.

- Parte II: A plataforma bioinformática que serve de base à parte II da sessão, designada HumanCyc -*Encyclopedia of Human Genes and Metabolism* (http://humancyc.org/) permitirá estudar uma biomolécula presente no metabolismo humano – colesterol -, tornando possível analisar de forma detalhada a via metabólica que lhe dá origem, reconhecendo ainda outros compostos da mesma e que desempenham um papel crucial no organismo (exemplo: na produção de vitamina D). Este conhecimento, para além do seu valor científico, ao estabelecer a relação entre a síntese de diferentes biomoléculas, poderá ser um mobilizador da adoção de comportamentos mais saudáveis. Pelas razões apontadas, este exercício tem um enquadramento curricular vasto, a saber: 9ºano: Saúde individual e comunitária (2;2.5); 10º ano: Módulo Inicial: Diversidade na biosfera – 2. A célula; 2.2. Constituintes básicos; 12ºano: Módulo 2 – Controlo de doenças e biotecnologia: 1.2. Desequilíbrios e doenças.

Sessão 6 (4h) – "A bioinformática ao serviço da população: exemplos práticos"

- Parte I: Na primeira parte desta sessão, utilizando a aplicação de análise evolutiva MEGA (*Molecular Evolutionary Genertics Analysis*) (http://www.megasoftware.net/), será feito um estudo sobre a evolução da resistência aos antibióticos (tetraciclina) em bactérias, que permita compreender melhor a emergência recente de estirpes de bactérias multirresistentes. O programa MEGA permite comparar diferentes organismos, analisando o fenómeno de evolução da resistência a antibióticos que se enquadra nas metas curriculares propostas para o 9ºano de

escolaridade no tema Saúde Individual e Comunitária: Resistência a antibióticos (1;1.6.). Serão ainda explorados os eventos de transferência vertical e lateral de genes.

- Parte II: Na segunda parte desta sessão, serão exploradas as ferramentas disponibilizadas no *Pathogen Modeling Program (PMP) Online* (http://pmp.errc.ars.usda.gov/PMPOnline.aspx) e no *ComBase* (http://www.combase.cc/index.php/en/) tendo como principais objetivos: identificar as valias de diferentes métodos de preservação de alimentos, e compreender o efeito de fatores extrínsecos (como a temperatura e concentração de oxigénio); de fatores intrínsecos (como o pH e osmolaridade); e de alguns aditivos alimentares. Esta atividade é ainda um ótimo paradigma para se compreender a importância de modelos preditivos para a indústria alimentar, sistematizados através de aplicações computacionais. Esta atividade tem a particularidade de se enquadrar em dois ciclos de ensino diferentes: 8ºano de escolaridade: Sustentabilidade na Terra – Gestão sustentável dos recursos – 18. Relacionar o desenvolvimento científico e tecnológico com a melhoria da qualidade de vida das populações humanas; e 12ºano: Módulo 2 – Controlo de Doenças e Biotecnologia: 3 – Microrganismos e Indústria Alimentar.

Sessão 7 (2,5h)

- Elaboração de uma proposta de exercício, enquadrado curricularmente, que permita potenciar e estender as ferramentas bioinformáticas utilizadas.

- Avaliação da ação de formação.

Metodologias

O curso de formação incluirá aulas de natureza teórico-prática e incidirá sobre atividades de resolução de exercícios, sempre que possível em grupos de dois formandos. Todas as atividades serão precedidas de uma introdução/fundamentação teórica e acompanhadas por instruções detalhadas num manual de apoio. Sempre que se justifique, será fornecida antecipadamente informação necessária e/ou adicional aos formandos.

Serão ainda considerados os seguintes procedimentos:

- Reflexão crítica sobre as atividades desenvolvidas, análise e discussão dos resultados obtidos;
- Discussão/reflexão sobre a natureza pedagógica de cada protocolo e a sua implementação em contexto escolar e de acordo com as orientações curriculares em vigor.

A abordagem de todos os conteúdos propostos para a ação de formação, será acompanhada de sumários e listas de bibliografia específica.

Condições de Frequência

Para que ocorra a avaliação da formação os formandos terão de cumprir um mínimo de dois terços do tempo previsto para as sessões presenciais (segundo o Regulamento para Acreditação e Creditação de Ações de Formação do Conselho Científico-Pedagógico de Formação Contínua).

Regime de avaliação dos formandos

Sem prejuízo de utilização de outras modalidades de avaliação complementares, a avaliação dos formandos será objecto de três tipos de avaliação:

- Avaliação contínua (40%): os formandos serão avaliados relativamente à participação nas atividades realizadas ao longo das sessões e pelo trabalho desenvolvido nas tarefas propostas pelo formador.
- Avaliação final (60%): no final da formação, a avaliação do aproveitamento dos formandos será feita através da produção de um trabalho escrito, nomeadamente de uma proposta de exercício que permita potenciar e estender as ferramentas bioinformáticas utilizadas ao longo do curso de formação.

Nos termos dos números 5 e 6 do artigo 4.º do mesmo Despacho, a avaliação a atribuir aos formandos é expressa numa classificação quantitativa na escala de 1 a 10 valores, tendo como referente as seguintes menções:

- Excelente — de 9 a 10 valores;
- Muito Bom — de 8 a 8,9 valores;
- Bom — de 6,5 a 7,9 valores;
- Regular — de 5 a 6,4 valores;
- Insuficiente — de 1 a 4,9 valores.

Referências

Amenkhienan, E., & Smith, E. (2006). A web-based genetic polymorphism learning approach for high school students and science teachers. *Biochem Mol Biol Educ*, 34(1):30-33. doi:10.1002/bmb.2006.49403401030

Boyle, J. (2004). Bioinformatics in undergraduate education: Practical examples. *Biochem Mol Biol Educ, 32*, 236-238. doi:10.1002/bmb.2004.494032040376

Caroline Alexandra, M. (2001). Building Phylogenetic Trees from DNA Sequence Data: Investigating Polar Bear & Giant Panda Ancestry. *The American Biology Teacher*, 63(9), 642-646. doi: 10.2307/4451210

Council, N. R. (2003). Engaging Schools: Fostering High School Students' Motivation to Learn. Washington, D.C.: The National Academies Press

Dudley, J. T., & Butte, A. J. (2009). A Quick Guide for Developing Effective Bioinformatics Programming Skills. *PLoS Comput Biol*, 5(12), e1000589. doi: 10.1371/journal.pcbi.1000589

Horta M., M. F., Nascimento R. (2012). Metas Curriculares - Tecnologias da Informação e da Comunicação 7º e 8ºanos: Ministério da Educação.

Ilomäki, L., Paavola, S., Lakkala, M., & Kantosalo, A. (2016). Digital competence – an emergent boundary concept for policy and educational research. *Education and Information Technologies*, 21(3), 655-679. doi: 10.1007/s10639-014-9346-4

Kovarik, D.N., Patterson, D.G., Cohen, C., Sanders, E.A., Peterson, K.A., Porter, S.G., Chowning, J.T. (2013). Bioinformatics education in high school: Implications for promoting science, technology, engineering, and mathematics careers. *CBE Life Sciences Education*, 12(3), 441-459. doi: 10.1187/cbe.12-11-0193

Machluf, Y., Gelbart, H., Ben-Dor, S., & Yarden, A. (2016). Making authentic science accessible-the benefits and challenges of integrating bioinformatics into a high-school science curriculum. *Brief Bioinform*. doi: 10.1093/bib/bbv113

Madigan M., M. J., Dunlap P., Clark D. (2010). Microbiologia de Brock (12nd ed.). Porto Alegre: artmed.

University, C. (2009). The Digital Literacy Project. Retrieved july, 2016, from https://digitalliteracy.cornell.edu/

Zylka, J., Christoph, G., Kroehne, U., Hartig, J., & Goldhammer, F. (2015). Moving beyond cognitive elements of ICT literacy: First evidence on the structure of ICT engagement. *Computers in Human Behavior*, 53, 149-160. doi:10.1016/j.chb.2015.07.008

Outras fontes consultadas

Bonito J., M. M., Silva M., Figueira D., Serrano M., Mesquita J., Rebelo H. (2013). Metas Curriculares - Ensino Básico 5.º, 6.º, 7.º e 8.º anos: Ministério da Educação

Bonito J., M. M., Silva M., Figueira D., Serrano M., Mesquita J., Rebelo H. (2014). Metas Curriculares - Ensino Básico - Ciências Naturais - 9º ano: Ministério da Educação e Ciência.

Mendes A., R. D. (2006). Programa de Biologia - 12ºano - Curso Científico-Humanístico de Ciências e Tecnologias: Ministério da Educação - Direção-Geral de Inovação e de Desenvolvimento Curricular.

Mendes A., R. D., Pinheiro E. (2001). Programa de Biologia e Geologia 10ºano - Componente de Biologia - Curso Científico-Humanístico de Ciências e Tecnologias: Ministério da Educação - Departamento do Ensino Secundário.

Mendes A., R. D., Pinheiro E. (2003). Programa de Biologia e Geologia 11º ano - Componente de Biologia - Curso Científico-Humanístico de Ciências e Tecnologias: Ministério da Educação - Departamento do Ensino Secundário.

Links úteis

http://blast.ncbi.nlm.nih.gov/Blast.cgi

http://humancyc.org/
(P. Romero, J. Wagg, M.L. Green, D. Kaiser, M. Krummenacker, and P.D. Karp Computational prediction of human metabolic pathways from the complete human genome, *Genome Biology* 6:R2 R2.1-17, 2004)

http://insilico.ehu.es/
(Bikandi, J., San Millán, R., Rementeria, A., and Garaizar, J. 2004. In silico analysis of complete bacterial genomes: PCR, AFLP-PCR, and endonuclease restriction. *Bioinformatics 20*:798-9. DOI: 10.1093/bioinformatics/btg491)

http://metacyc.org/
(Caspi, R. Billington. R, Ferrer, L. Fulcher. C.A, Keseler. I.M, Kothari. A, Krummenacker. M, Latendresse. M, Mueller. LA, Ong. Q, Paley. S, Subhraveti. P, Weaver. DS, and Karp, P.D.(2014) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases Nucleic Acids Research 44(1):D471-80)

http://pmp.errc.ars.usda.gov/PMPOnline.aspx

http://www.ncbi.nlm.nih.gov/

http://www.ncbi.nlm.nih.gov/orffinder/

https://edgar.computational.bio.uni-giessen.de/

https://genomevolution.org/CoGe/
(Eric Lyons, Michael Freeling (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences *The Plant Journal 53* (4) , 661-673)

https://www.genoscope.cns.fr/agc/microscope/home/index.php

Biologia Molecular: Análises *in Silico* (Parte I)

Recorrendo à plataforma bioinformática *In silico simulation of molecular biology experiments* (http://insilico.ehu.es/) será feita a análise das enzimas de restrição e o seu funcionamento, assim como a simulação *in silico* de experiências com recurso à técnica de PCR. Apesar das potencialidades desta plataforma serem mais alargadas, será dado ênfase a duas aplicações que vão de encontro a tópicos curriculares atuais: i) Amplificação por PCR, tendo como objetivo a reconstituição e confirmação *in silico* de uma experiência de diagnóstico de bactérias patogénicas reportada num artigo científico; e ii) A elaboração de mapas de restrição de genomas de diferentes estirpes bacterianas com diferentes enzimas de restrição.

Enquadramento curricular:

As componentes desta abordagem vão ao encontro das orientações curriculares do programa de Biologia do 12ºano, essencialmente materializadas no tema 4 - Alterações do Material Genético: 4.2. Fundamentos da Engenharia Genética, em que são abordadas matérias como a importância das enzimas de restrição como ferramentas da engenharia genética.

Exercício 1: *Determinar a especificidade de primers para deteção de S. aureus*

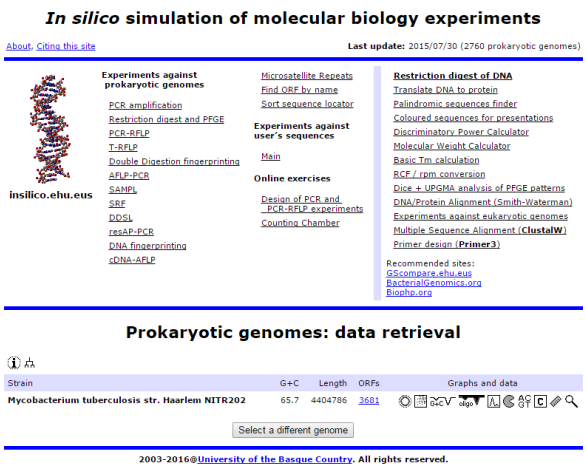
1.1. Identificar no excerto do artigo abaixo fornecido os *primers* usados para deteção *S. aureus*.

Table 4
Amplification primers of the multiplex PCR.

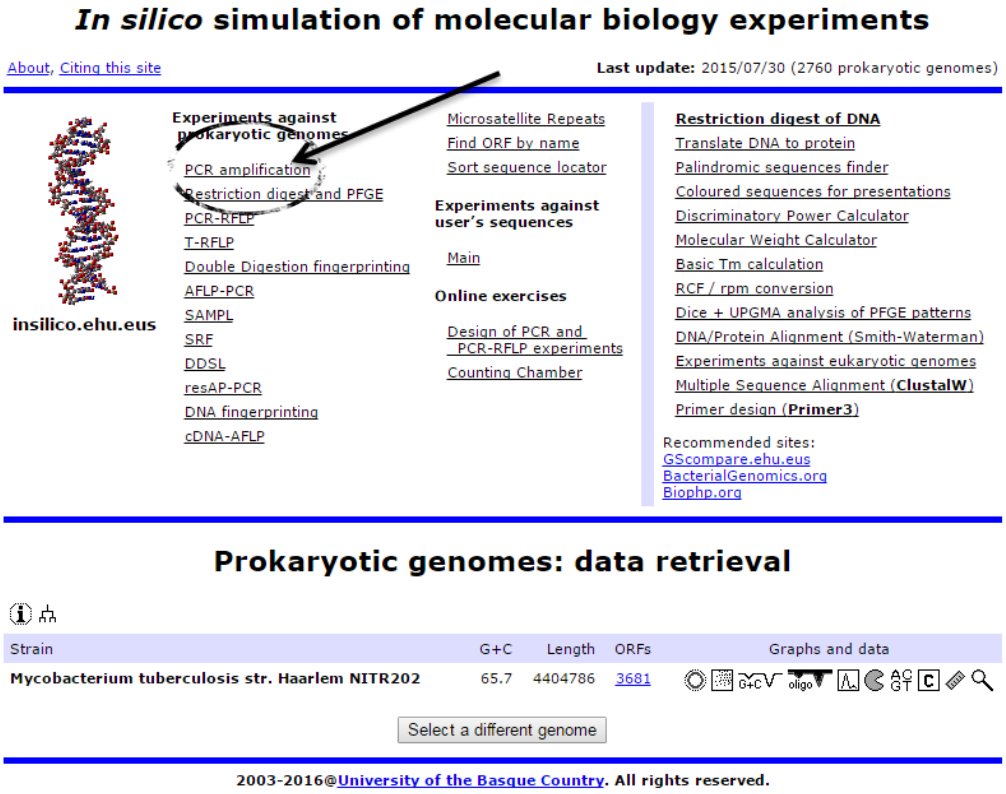
Strains	Amplification gene	Primer	Primer sequences (5'-3')	Amplified fragment
Salmonella spp.	xcd	sc8	ATCGTGATACAGAACGCCG TCTTCGTCATCCACCCAGA	537 bp
S. aureus	Vick	sa5	AAAGGTGTAGGTTGGAAAGTAGAAG GTTACAGGCATTTTGTCTTTAGGTG	176 bp
L. monocytogenes	LMOf 2365-2721	lm16	CTGTTCGTCGGTCCGTGGTA CAATGCTCTAATGCGGTGGT	1623 bp

(Yu, Q., Zhai, L., Bie, X., Lu, Z., Zhang, C., Tao, T., Zhao, H. (2016). Survey of five food-borne pathogens in commercial cold food dishes and their detection by multiplex PCR. *Food Control, 59*, 862-869.)

1.2. Aceder ao link: http://insilico.ehu.es/



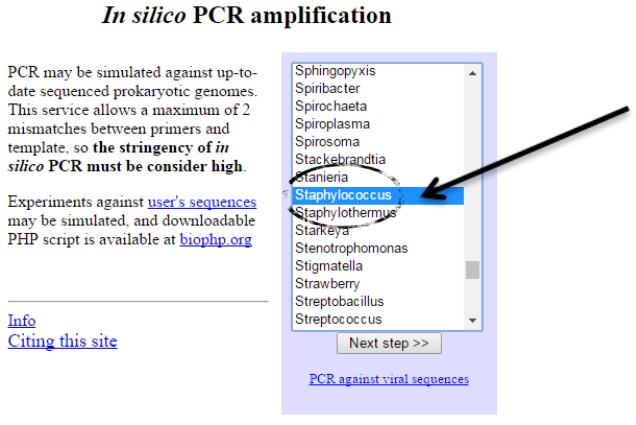
1.3. Selecionar ***PCR amplification*** (Amplificação por PCR).



1.4. Escolher o género bacteriano *Staphylococcus*, que se pretendeu identificar no estudo.

Table 4
Amplification primers of the multiplex PCR.

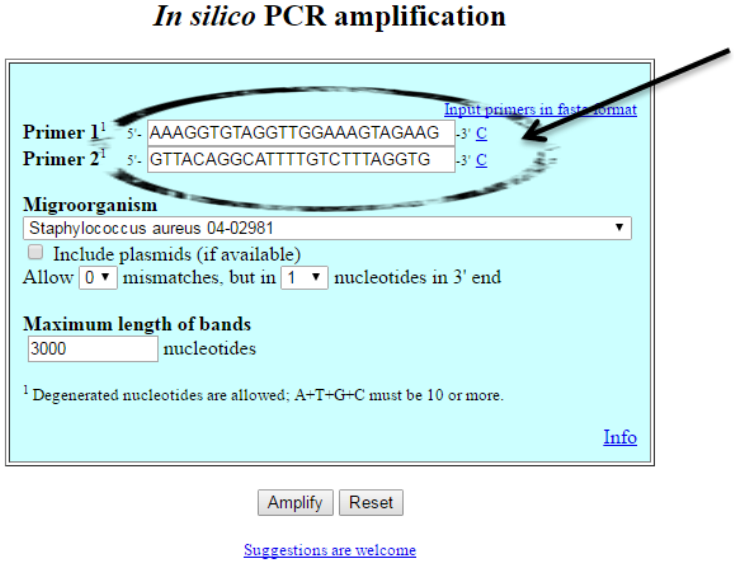
Strains	Amplification gene	Primer	Primer sequences (5′–3′)	Amplified fragment
Salmonella spp.	xcd	sc8	ATCGTGATACAGAACGCCG TCTTCGTCATCCACCCAGA	537 bp
S. aureus	Vick	sa5	AAAGGTGTAGGTTGGAAAGTAGAAG GTTACAGGCATTTTGTCTTTAGGTG	176 bp
L. monocytogenes	LMOf 2365-2721	lm16	CTGTTCGTCGGTCCGTGGTA CAATGCTCTAATGCGGTGGT	1623 bp



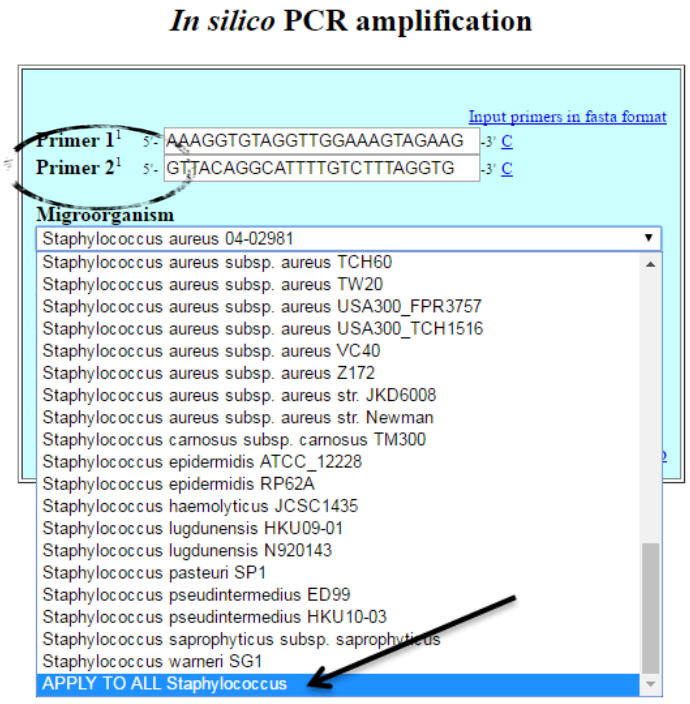
1.5. Inserir os *primers* utilizados no estudo para deteção desta bactéria.

Table 4
Amplification primers of the multiplex PCR.

Strains	Amplification gene	Primer	Primer sequences (5'-3')	Amplified fragment
Salmonella spp.	xcd	sc8	ATCGTGATACAGAACGCCG TCTTCGTCATCCACCCAGA	537 bp
S. aureus	Vick	sa5	AAAGGTGTAGGTTGGAAAGTAGAAG GTTACAGGCATTTTGTCTTTAGGTG	176 bp
L. monocytogenes	LMOf 2365-2721	lm16	CTGTTCGTCGGTCCGTGGTA CAATGCTCTAATGCGGTGGT	1623 bp

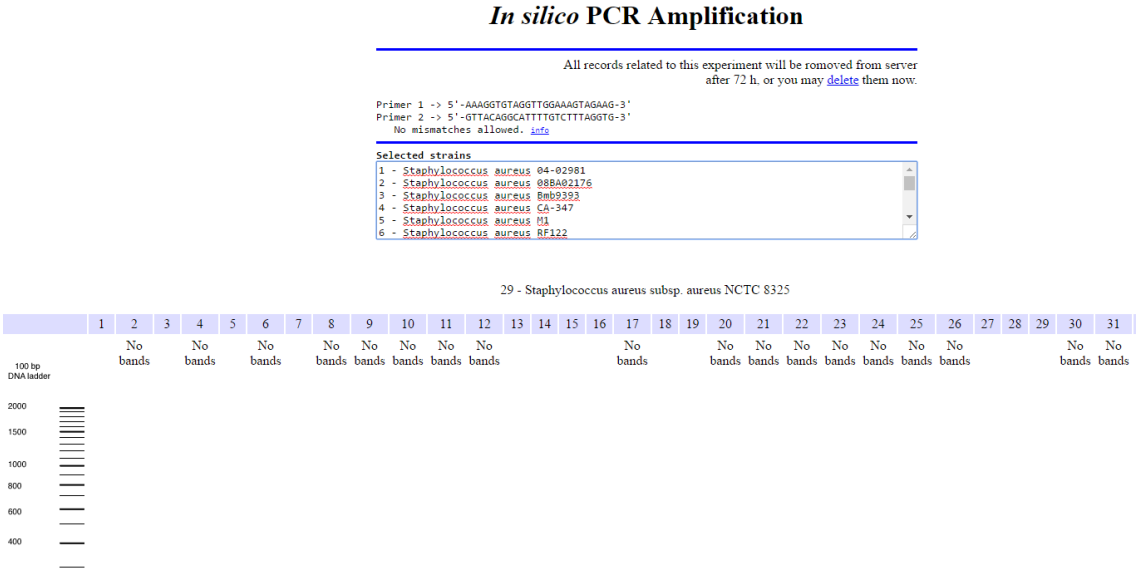


1.6. Selecionar **Apply to all *Staphylococcus*** (Aplicar a todos os *Staphylococcus)*

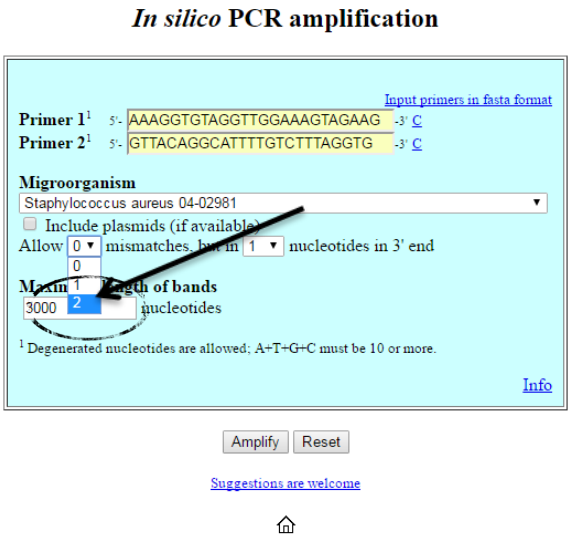


1.7. Clicar em ***Amplify*** (amplificar).

1.8. Concluir acerca da especificidade dos *primers* utilizados para identificação da bactéria em estudo.

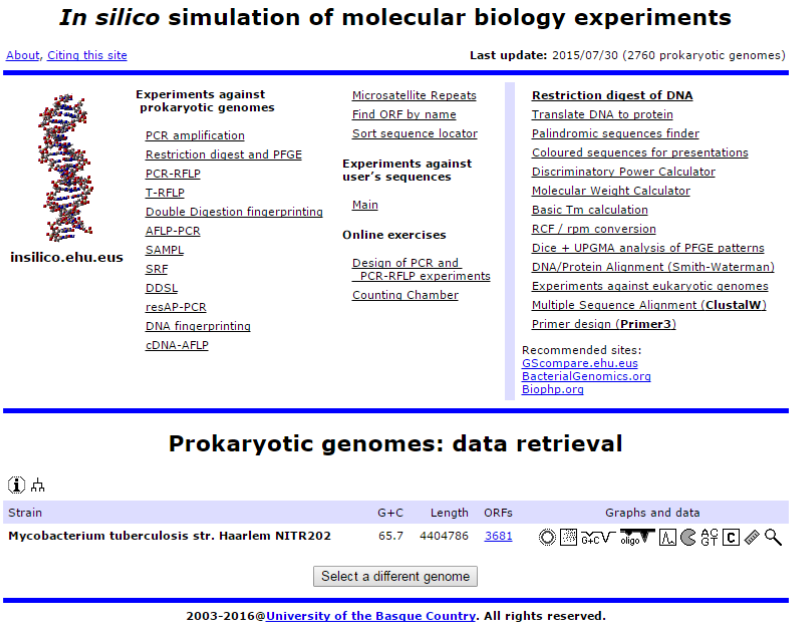


1.9. Repetir os passos **1.5.** a **1.8.** permitindo 2 *mismatches*.

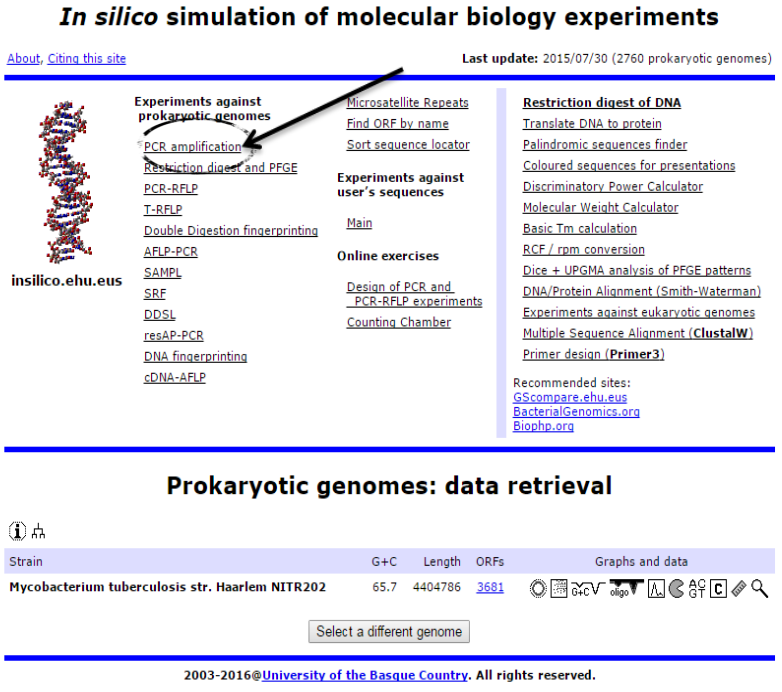


Exercício 2: *Determinar a especificidade de primers para deteção de Salmonella spp.*

2.1. Aceder ao link: http://insilico.ehu.es/



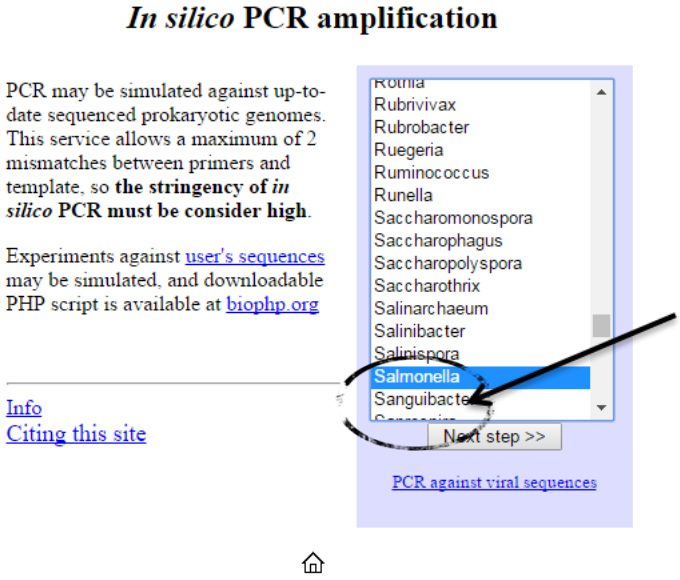
2.2. Selecionar ***PCR amplification*** (amplificação por PCR).



2.3. Escolher o género bacteriano *Salmonella*, que se pretendeu identificar no estudo.

Table 4
Amplification primers of the multiplex PCR.

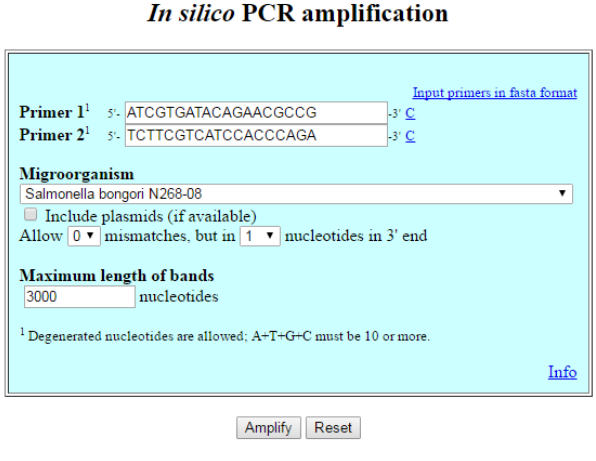
Strains	Amplification gene	Primer	Primer sequences (5′–3′)	Amplified fragment
Salmonella spp.	xcd	sc8	ATCGTGATACAGAACGCCG TCTTCGTCATCCACCCAGA	537 bp
S. aureus	Vick	sa5	AAAGGTGTAGGTTGGAAAGTAGAAG GTTACAGGCATTTTGTCTTTAGGTG	176 bp
L. monocytogenes	LMOf 2365-2721	lm16	CTGTTCGTCGGTCCGTGGTA CAATGCTCTAATGCGGTGGT	1623 bp



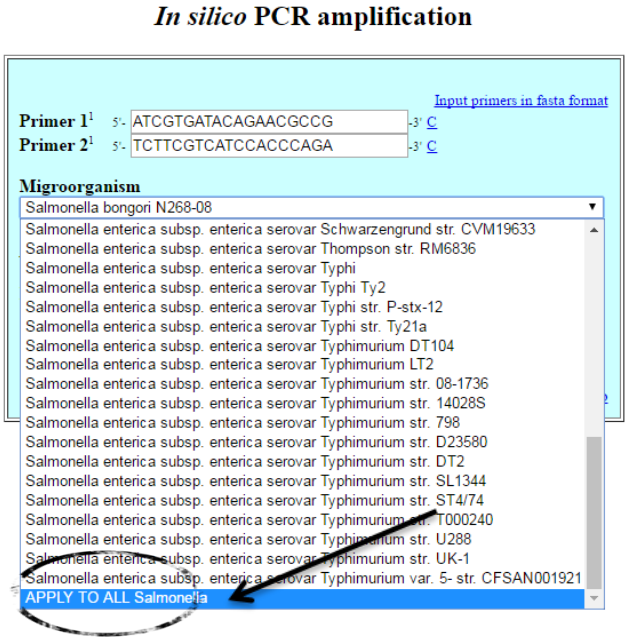
2.4. Inserir os *primers* utilizados no estudo para deteção desta bactéria.

Table 4
Amplification primers of the multiplex PCR.

Strains	Amplification gene	Primer	Primer sequences (5′–3′)	Amplified fragment
Salmonella spp.	xcd	sc8	ATCGTGATACAGAACGCCG TCTTCGTCATCCACCCAGA	537 bp
S. aureus	Vick	sa5	AAAGGTGTAGGTTGGAAAGTAGAAG GTTACAGGCATTTTGTCTTTAGGTG	176 bp
L. monocytogenes	LMOf 2365-2721	lm16	CTGTTCGTCGGTCCGTGGTA CAATGCTCTAATGCGGTGGT	1623 bp

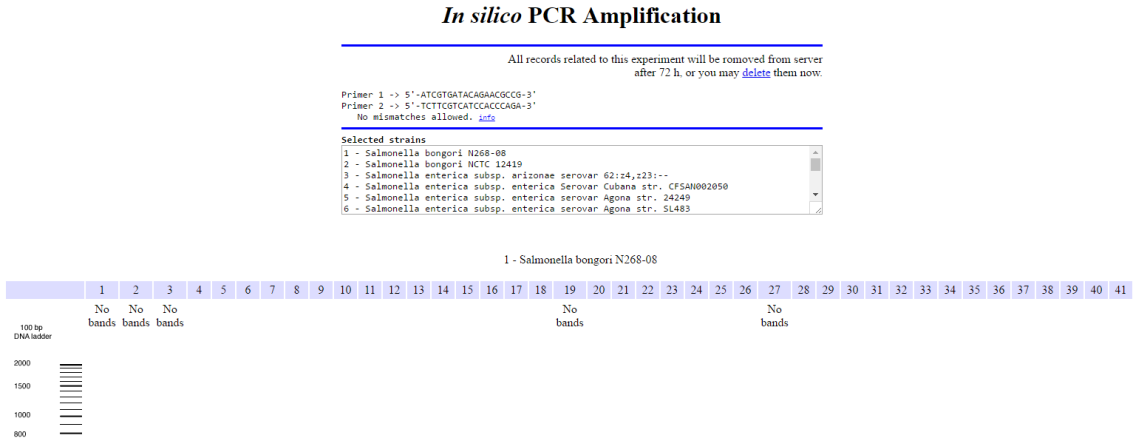


2.5. Selecionar: ***Apply to all Salmonella*** (Aplicar a todas as *Salmonella*)

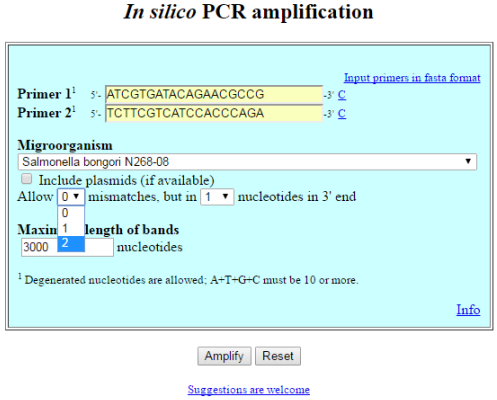


2.6. Clicar em ***Amplify*** (amplificar)

2.7. Concluir acerca da especificidade dos *primers* utilizados para identificação da bactéria em estudo.

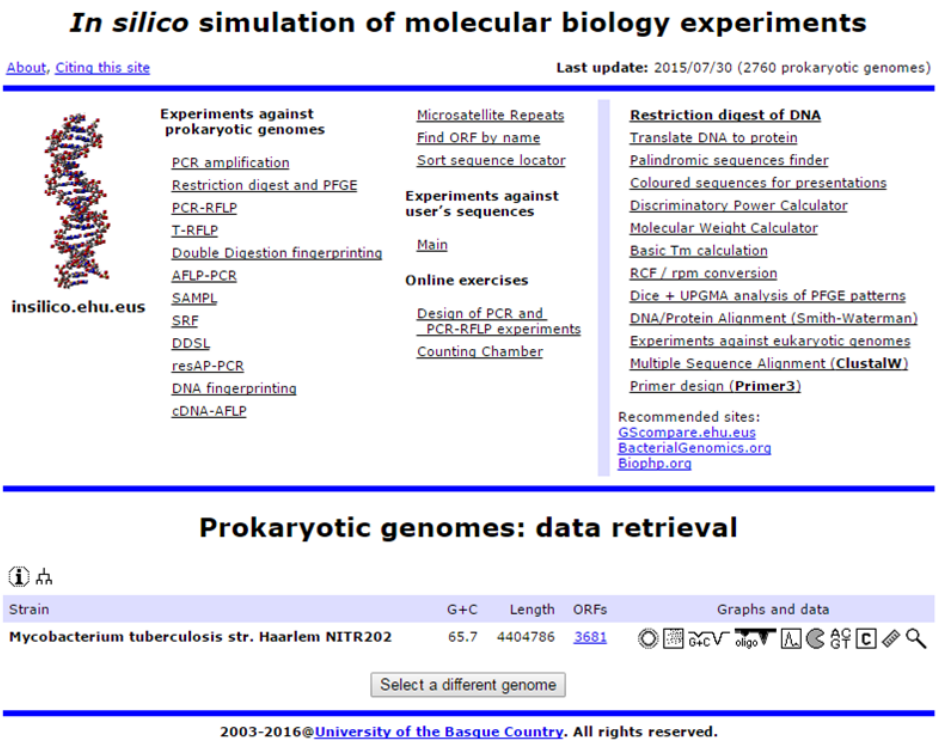


2.8. Repetir os passos **2.4.** a **2.7.** permitindo 2 *mismatches*.

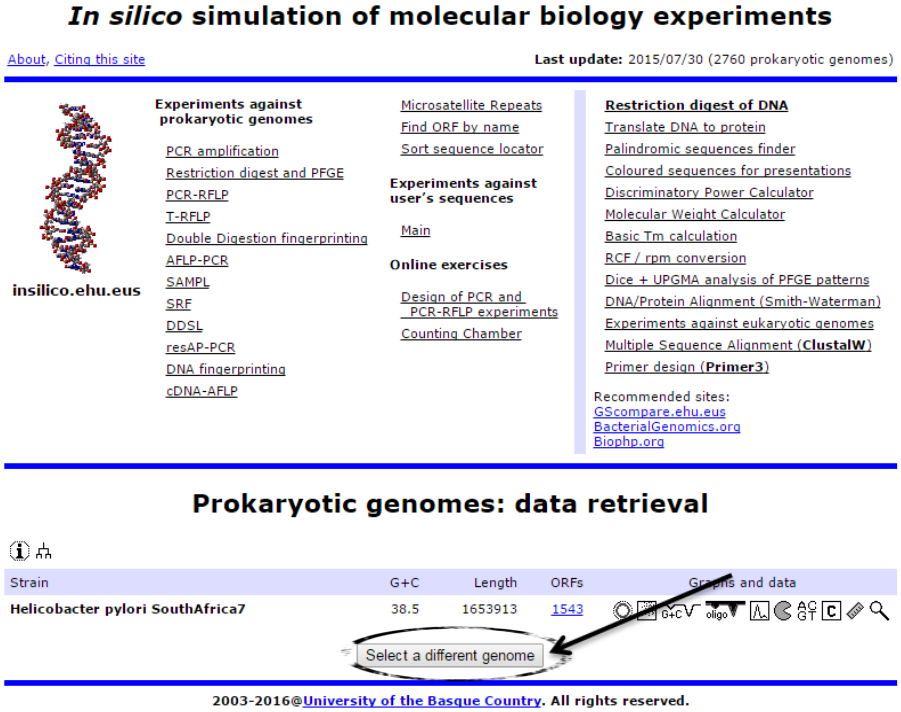


Exercício 3: *Elaborar mapas de restrição*

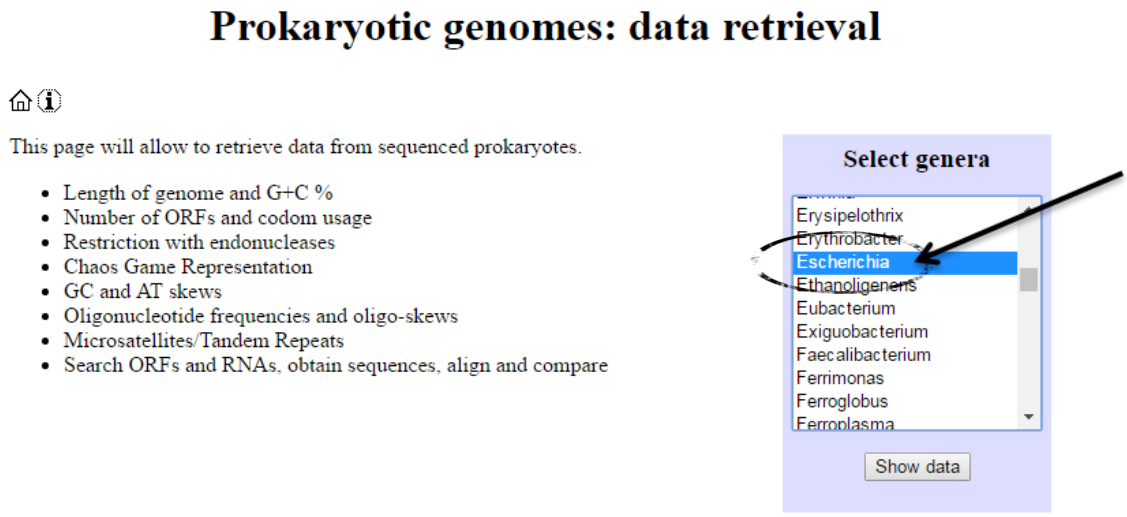
3.1. Aceder ao link: http://insilico.ehu.es/



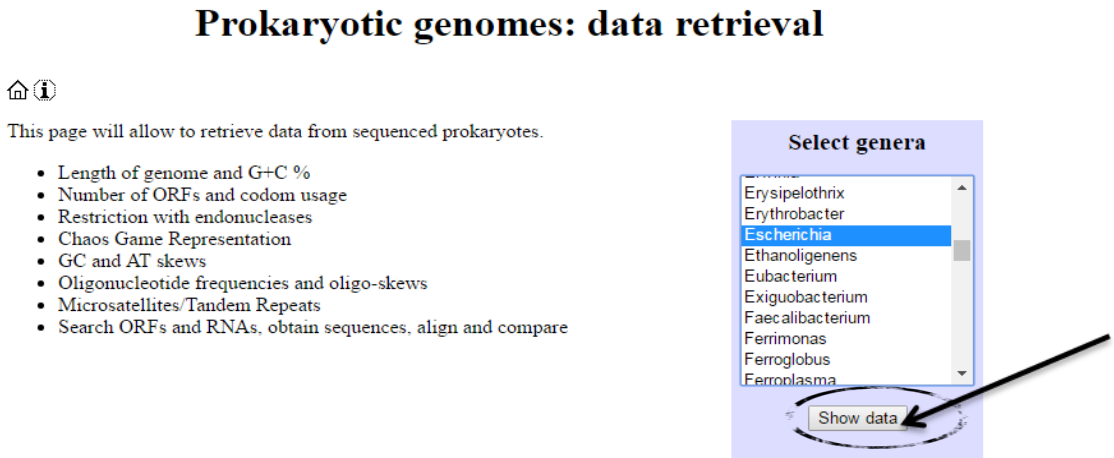
3.2. No final da página selecionar: ***Select a diferent genome*** (Selecionar um genoma diferente).



3.3. Escolher o género *Escherichia.*



3.4. Clicar em ***Show Data*** (mostrar dados)



3.5. Selecionar ***Escherichia coli* 0127:H6 E2348/69**, escolhendo a opção de gráfico assinalada em baixo:

3.6. Identificar a sequência reconhecida pela enzima Abs I.

3.7. Clicar no número de fragmentos originados pela enzima Abs I.

Restriction digest of Escherichia coli 0127:H6 E2348/69

This service shows theoretical number of cuts generated by commercial restriction enzymes.
When the number of cuts is less than 50, PFGE simulation is available by following the link.

Filter results

Restriction enzymes	Sequence	Fragments
AarI	CACCTGCNNNN'NNNN_	1140
AasI,DrdI,DseDI	GACNN_NN'NNGTC	759
AatI,Eco147I,PceI,SseBI,StuI	AGG'CCT	531
AatII	G_ACGT'C	758
AbsI	CC'TCGA_GG	6
Acc16I,AviII,FspI,NsbI	TGC'GCA	2155
Acc36I,BfuAI,BspMI,BveI	ACCTGCNNNN'NNNN_	4019
Acc65I,Asp718I	G'GTAC_C	548
AccB1I,BanI,BshNI,BspT107I	G'GYRC_C	3845
AccB7I,BasI,PflMI,Van91I	CCAN_NNN'NTGG	1718
AccBSI,BsrBI,MbiI	CCG'CTC	2001
AccI,FblI,XmiI	GT'MK_AC	1676

3.7. Clicar no número de fragmentos originados pela enzima Abs I.

3.8. Analisar as características dos fragmentos gerados.

***In silico* restriction map and PFGE**

Genome: Escherichia coli 0127:H6 E2348/69
Length of genome: 4965553 bp
Restriction Enzyme: AbsI
Recognition sequence: CC'TCGA_GG
Number of Cleaves: 6

Simulate PFGE for all Escherichia
info

Cleavage Position	Length of sequence	Length of sequence (**sorted**)	PFGE
506094	506101	1319656	
1148815	642721	1237722	
2386537	1237722	903863	727.5 kb
3706193	1319656	642721	
4061683	355490	506101	485.5 kb
4965546	903863	355490	388.0 kb

3.9. Clicar em *all Escherichia*

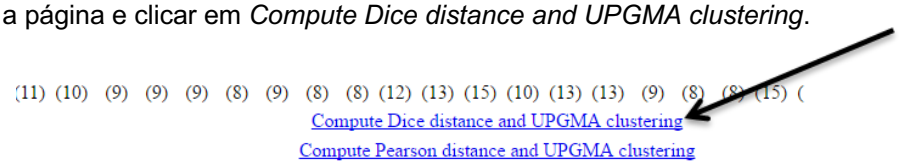
***In silico* restriction map and PFGE**

Genome: Escherichia coli 0127:H6 E2348/69
Length of genome: 4965553 bp
Restriction Enzyme: AbsI
Recognition sequence: CC'TCGA_GG
Number of Cleaves: 6

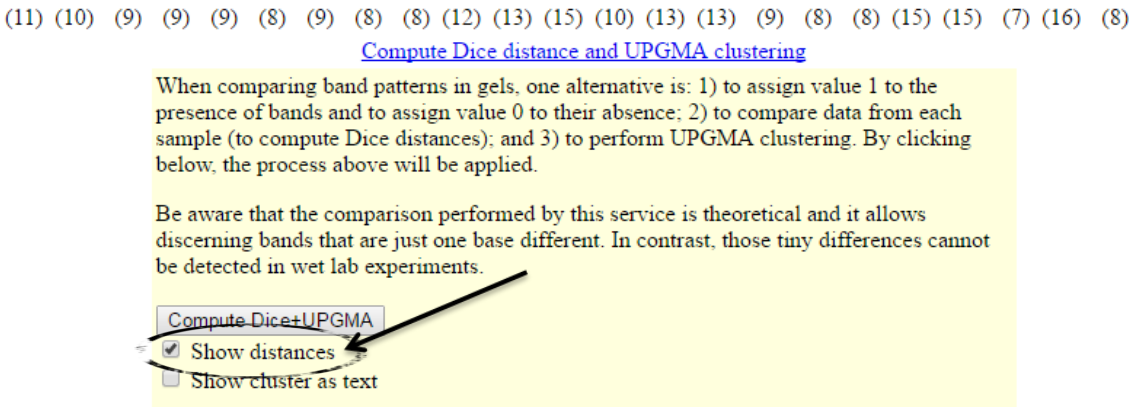
Simulate PFGE for all Escherichia
info

Cleavage Position	Length of sequence	Length of sequence (**sorted**)	PFGE
506094	506101	1319656	
1148815	642721	1237722	
2386537	1237722	903863	727.5 kb

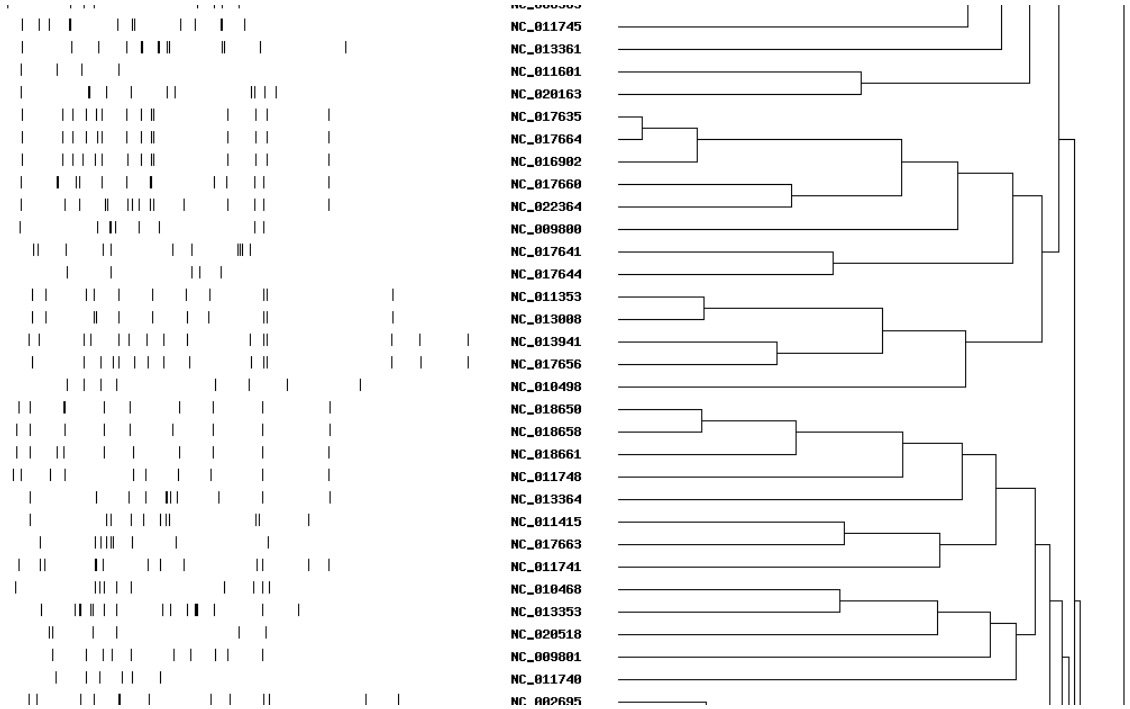
3.10. Descer a página e clicar em *Compute Dice distance and UPGMA clustering*.



3.11. Clicar em *Show distances e depois acionar o botão Compute Dice+UPGMA*.



3.12. Analisar os resultados obtidos.



*Ao comparar padrões de banda em géis, uma alternativa é: atribuir o valor 1 à presença de bandas e atribuir o valor 0 à sua ausência; Comparar os dados de cada amostra (para calcular ***Dice distances***). Uma outra alternativa é determinar uma matriz de dados para cada linha composta da posição no géis (distância), e cor nessa posição, e comparar os *arrays* anteriores (para calcular ***Pearson's distances***).

Biologia Molecular: Análises *in Silico* (Parte II)

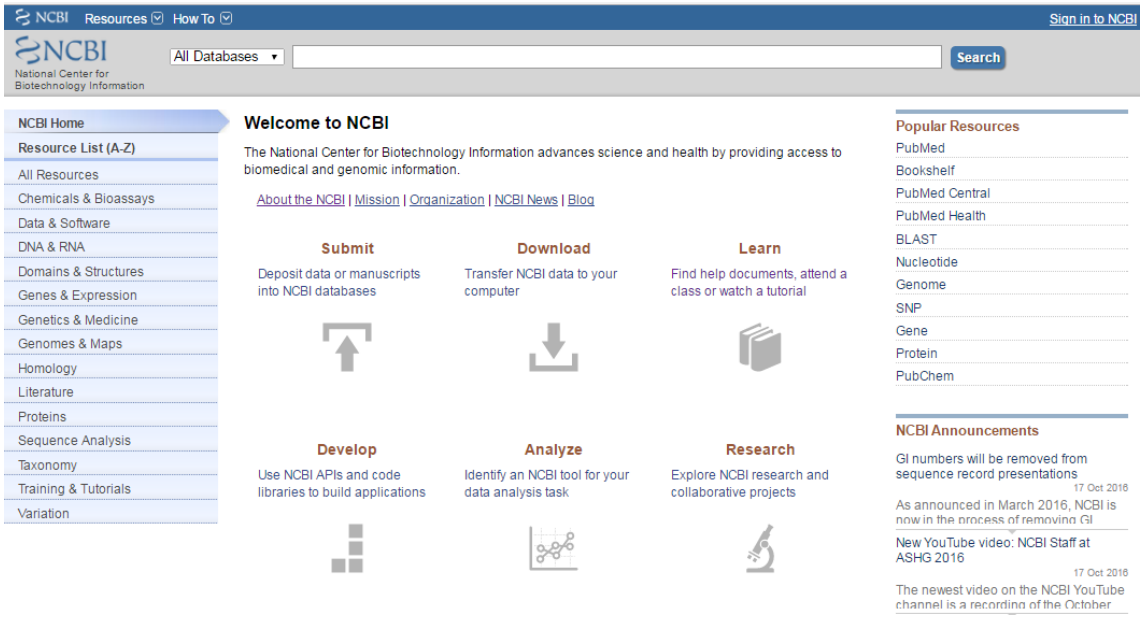
Serão exploradas ferramentas de análise de sequências de DNA. Recursos bioinformáticos, nomeadamente o *NCBI ORF finder* (http://www.ncbi.nlm.nih.gov/orffinder/) e o chamado BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), serão usados para análise *in silico* de um fragmento de DNA cuja função é desconhecida. Será possível identificar os genes putativos e suas funções, comparando com sequências já conhecidas. Este conhecimento contribuirá para uma abordagem holística de noções básicas como genoma, gene e a importância de regiões intergénicas.

Enquadramento curricular:

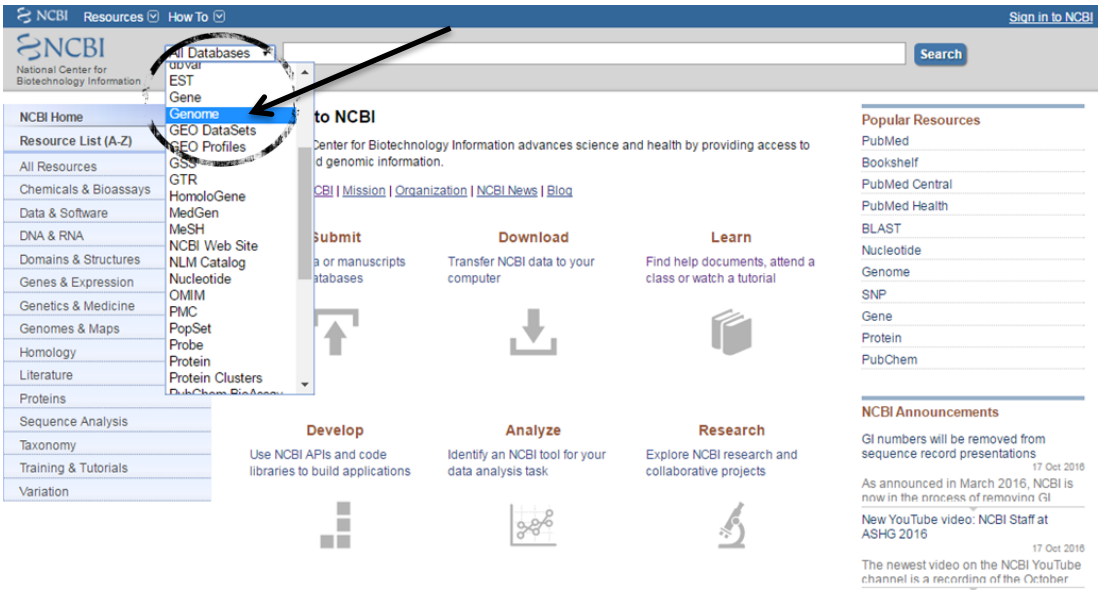
Este exercício permitirá explorar ao nível do 12ºano de escolaridade a organização do material genético, enfatizando, por exemplo, os cromossomas como entidades que contêm genes ou o papel dos operões para os seres procariontes e dos componentes intervenientes em mecanismos de regulação (Módulo 1: Reprodução e Património Genético – 3.2. Organização e regulação do material genético).

Exercício 4: *Aceder a uma sequência de DNA*

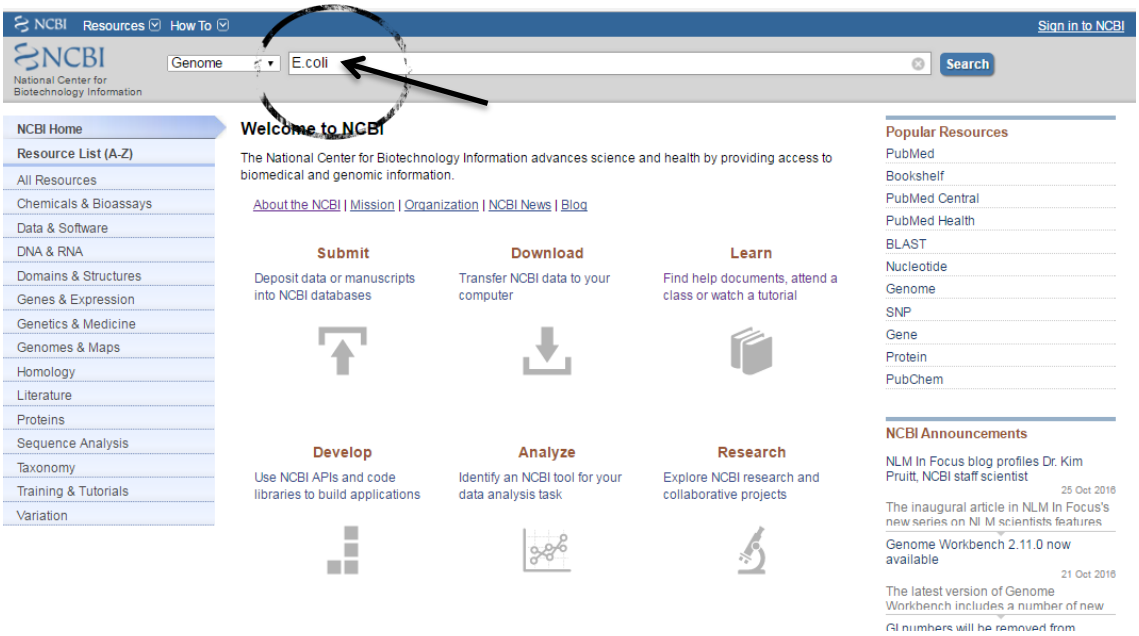
4.1. Aceder à plataforma NCBI através do link: http://www.ncbi.nlm.nih.gov/



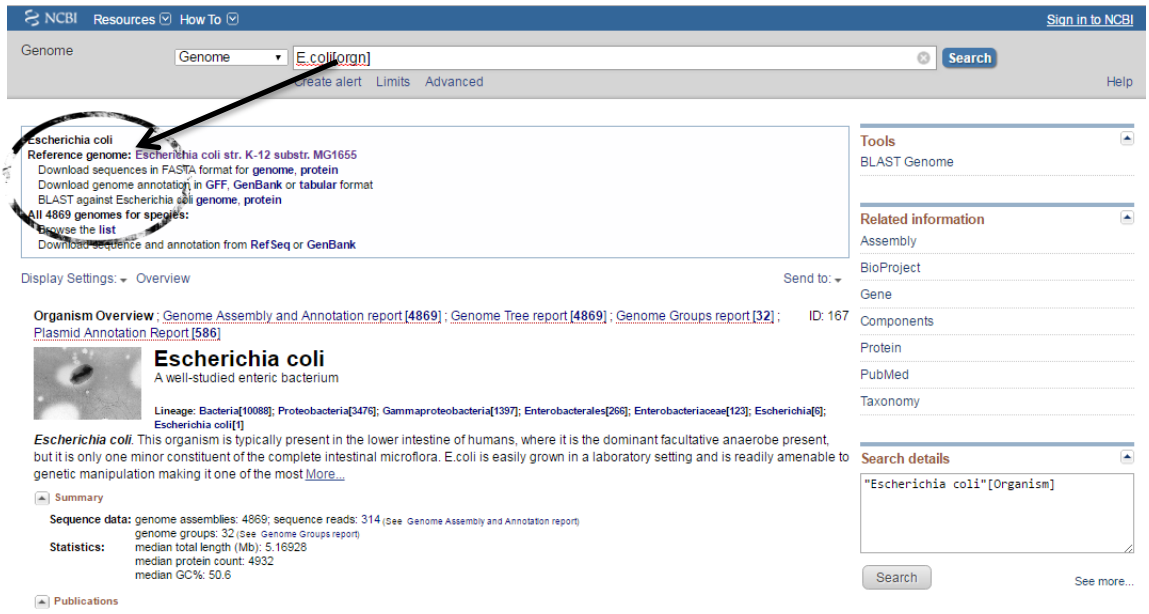
4.2. No menu de busca, selecionar: ***Genome*** (Genoma)



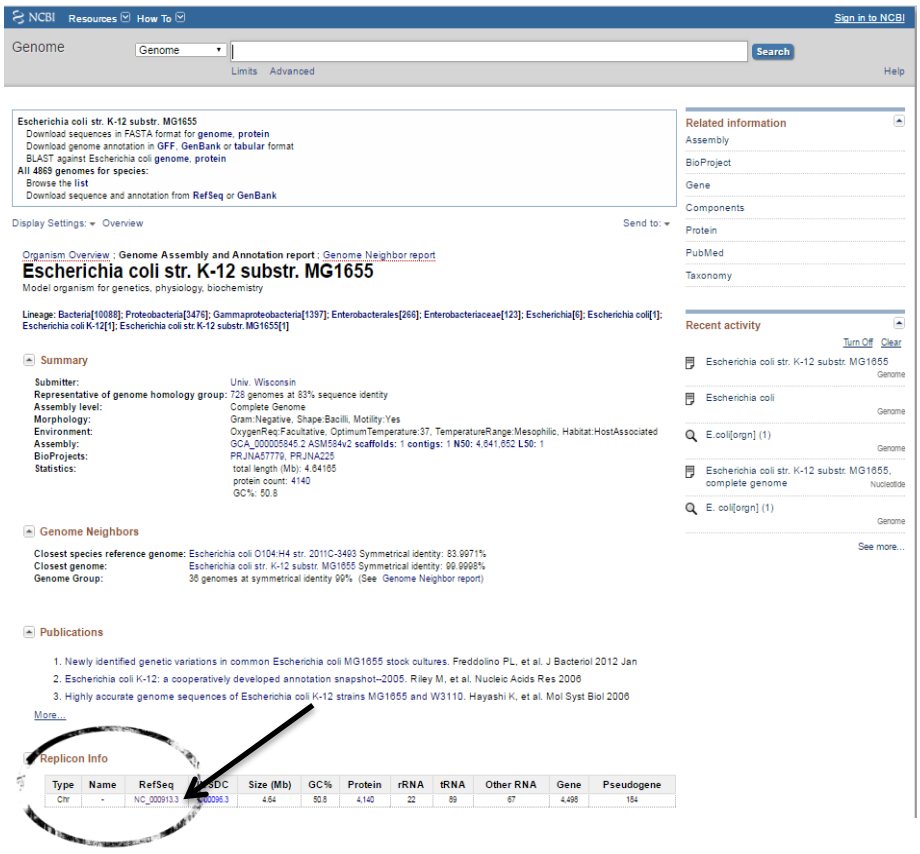
4.3. Procurar por "***E.coli***" e clicar em ***Search*** (pesquisar)



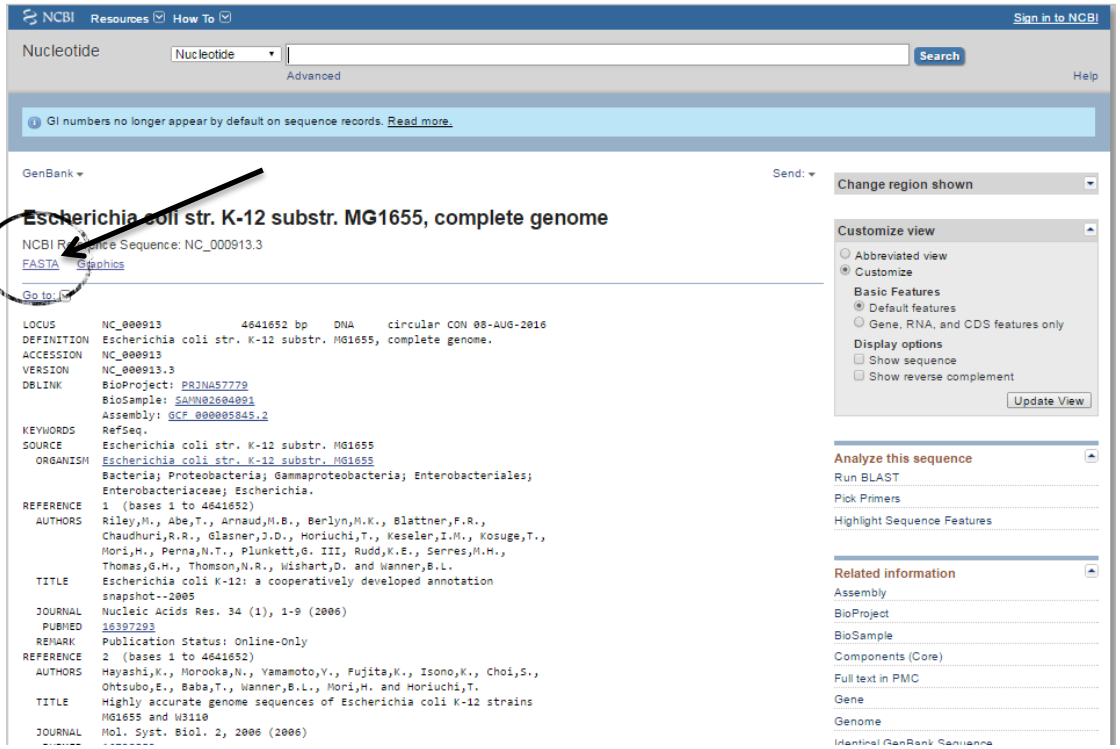
4.4. No início da nova página, selecionar o ***Reference Genome*** (Genoma de Referência)



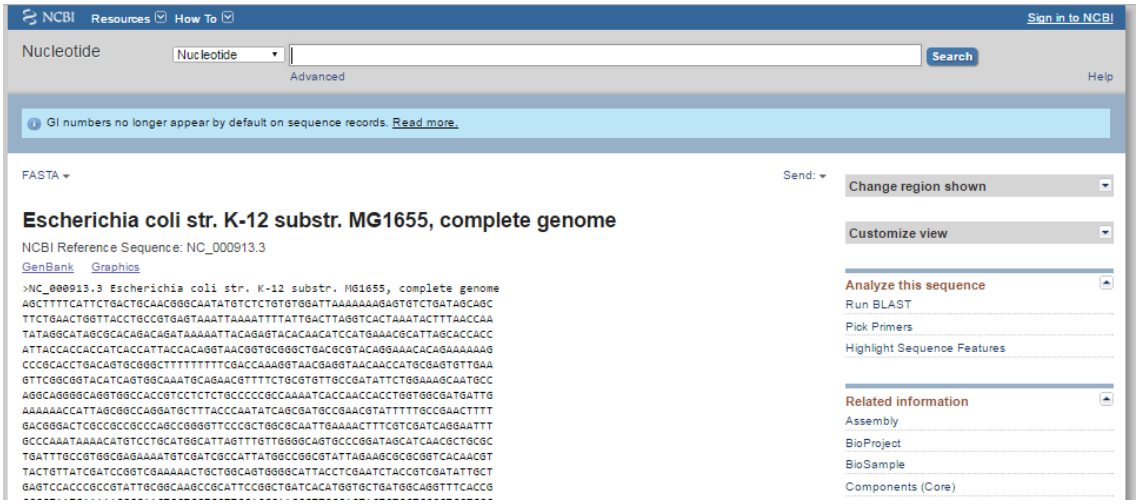
4.5. Descer a página até encontrar o campo: ***Replicon Info*** e clicar em **NC_000913.3**

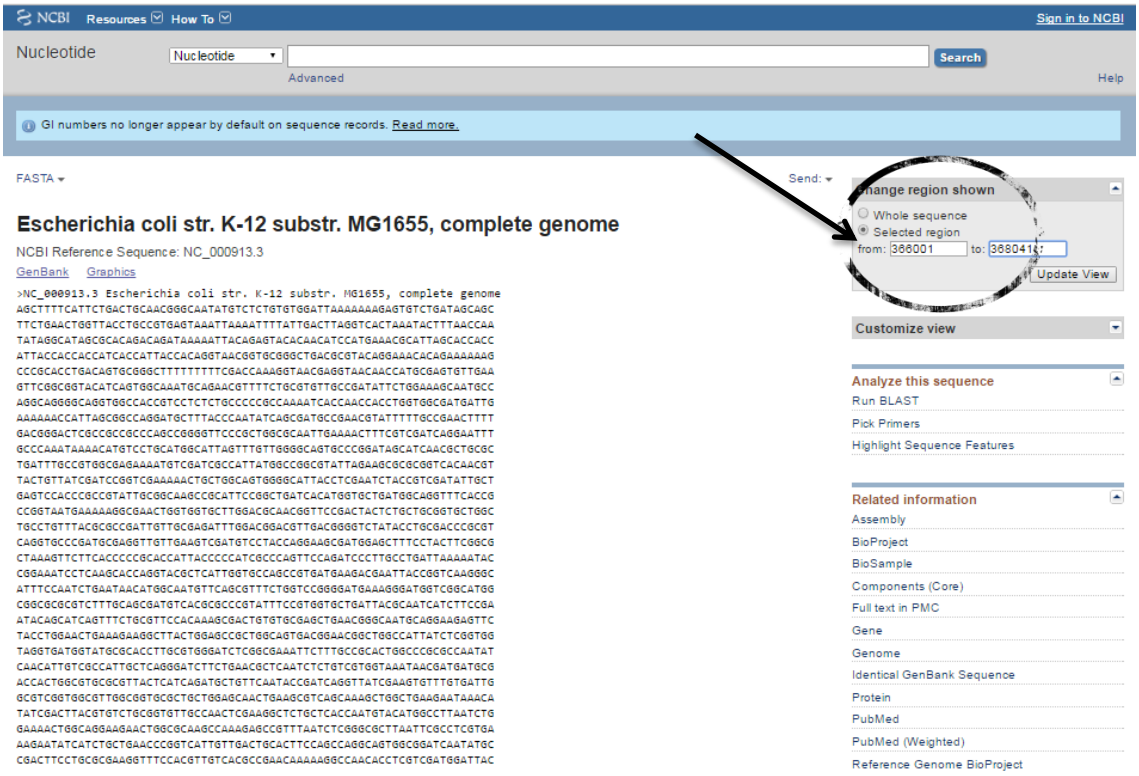


4.6. Escolher a opção de **formato FASTA**

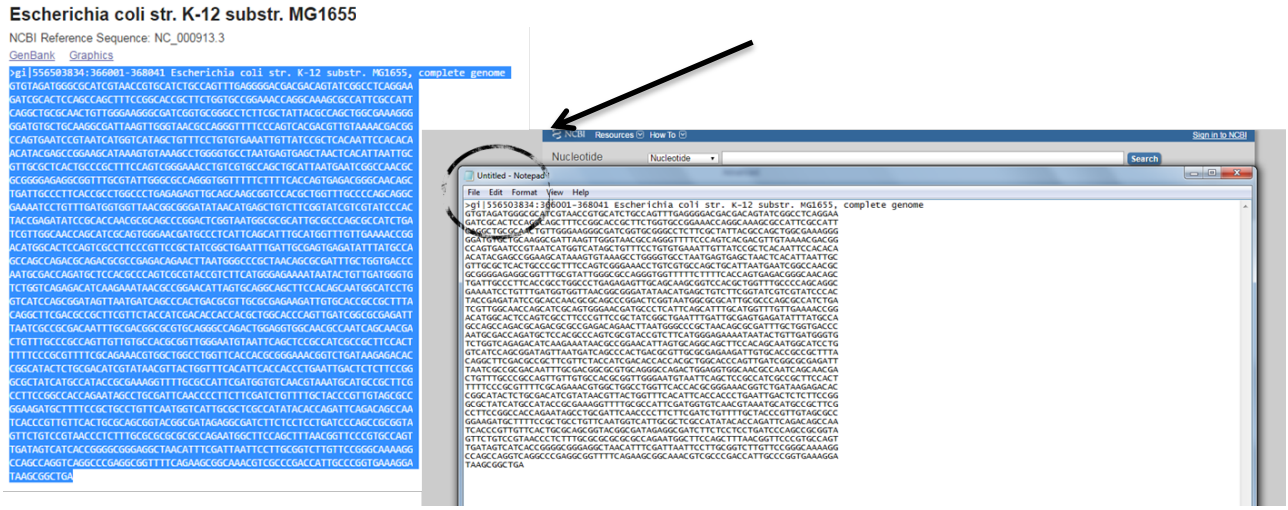


4.7. Abrir a caixa de seleção ***Change region shown*** (alterar região a apresentar) e inserir as coordenadas: **366001-368041**



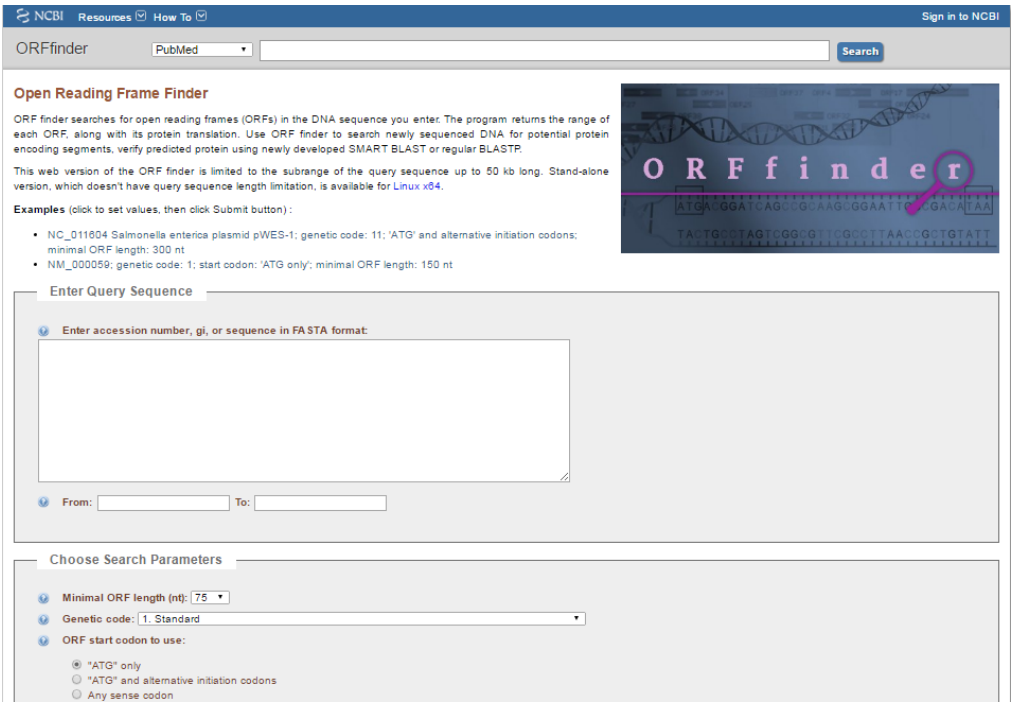


4.8. Copiar e colocar toda a sequência num editor de texto (exemplo: Bloco de Notas)

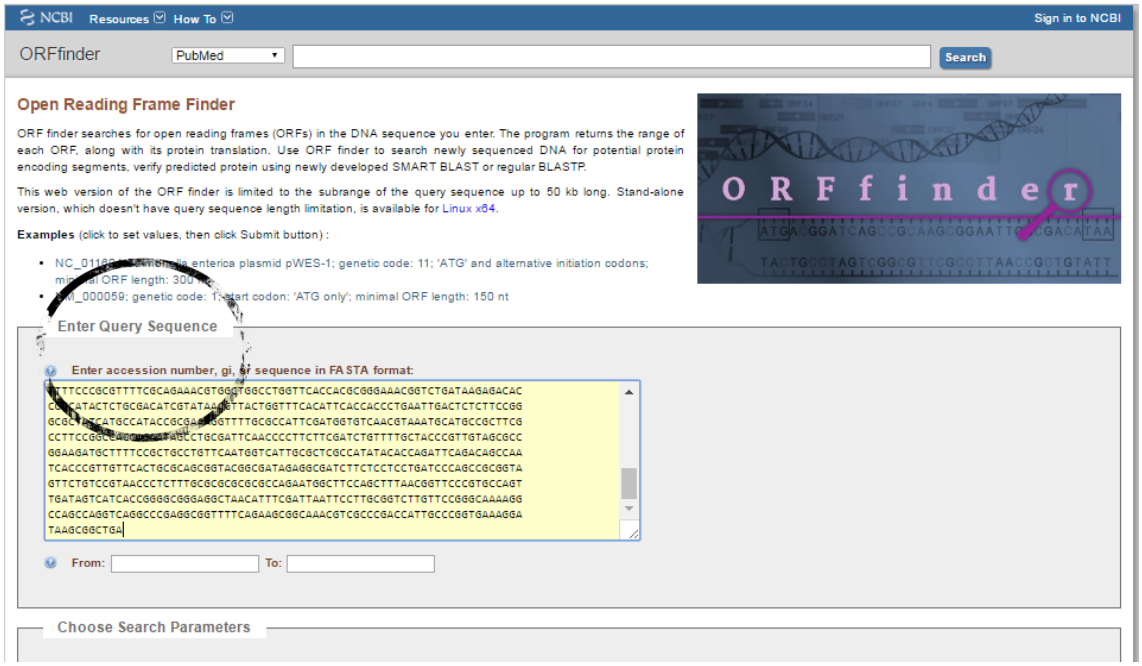


Exercício 5: *Identificação de ORF's (codões de iniciação alternativos)*

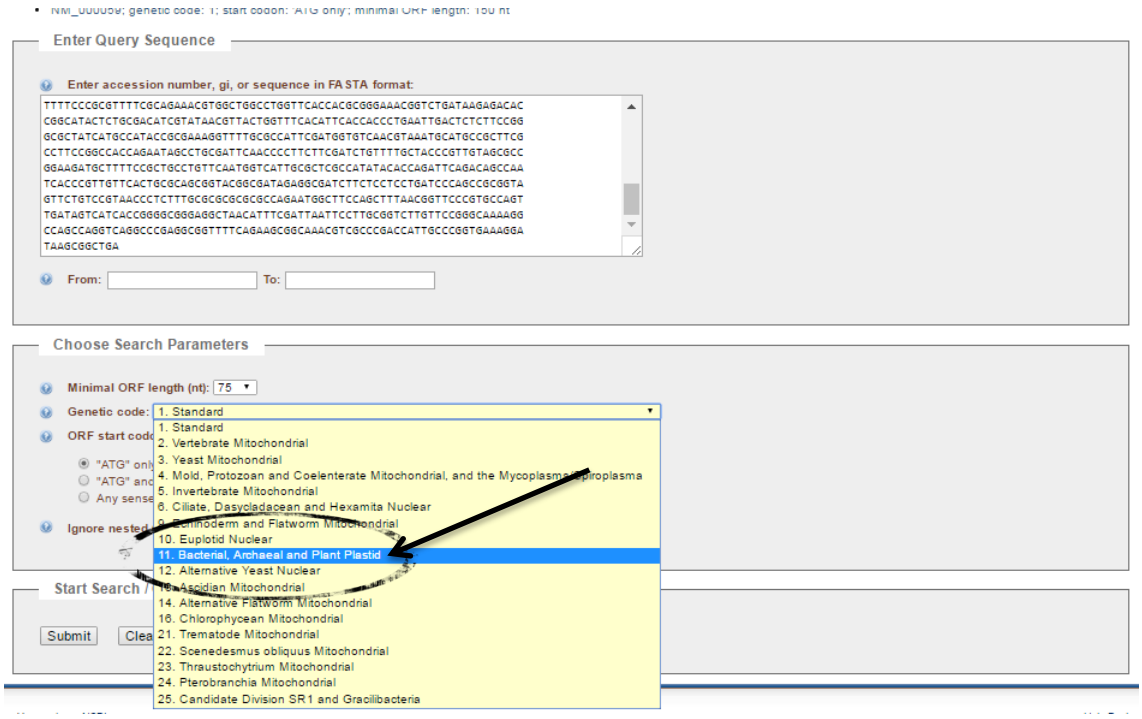
5.1. Aceder à página "*ORFfinder*" da plataforma NCBI através do link: http://www.ncbi.nlm.nih.gov/orffinder/



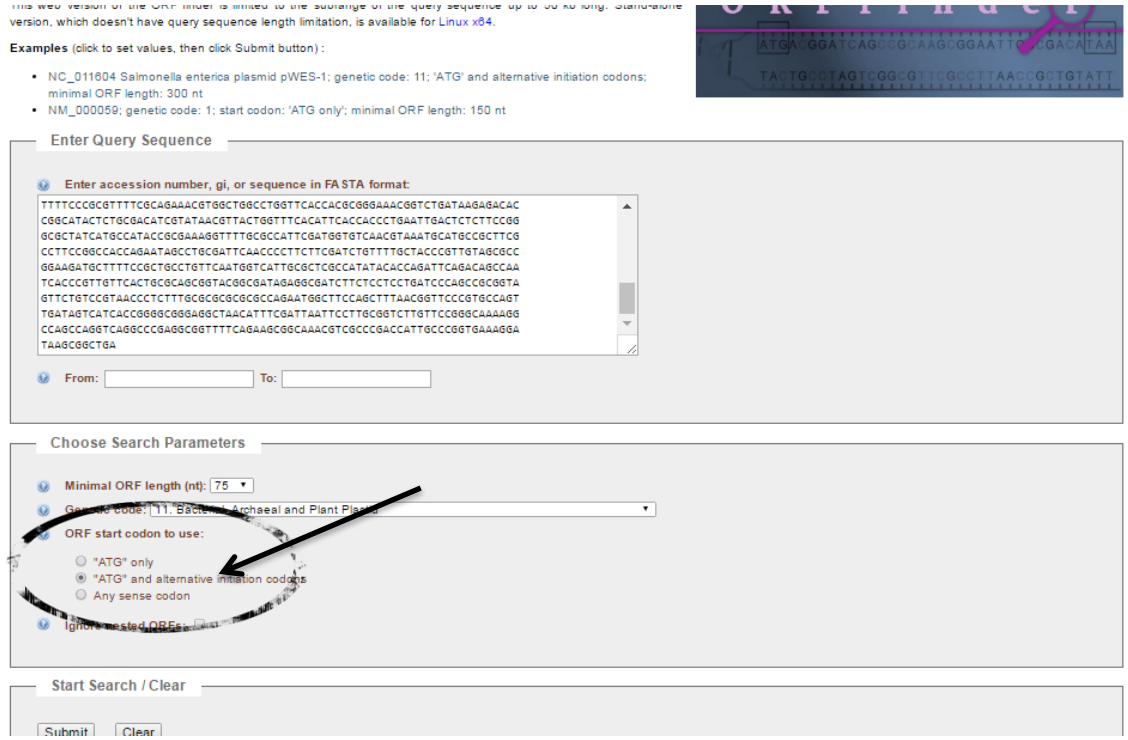
5.2. Inserir a sequência obtida previamente e anotada no processador de texto



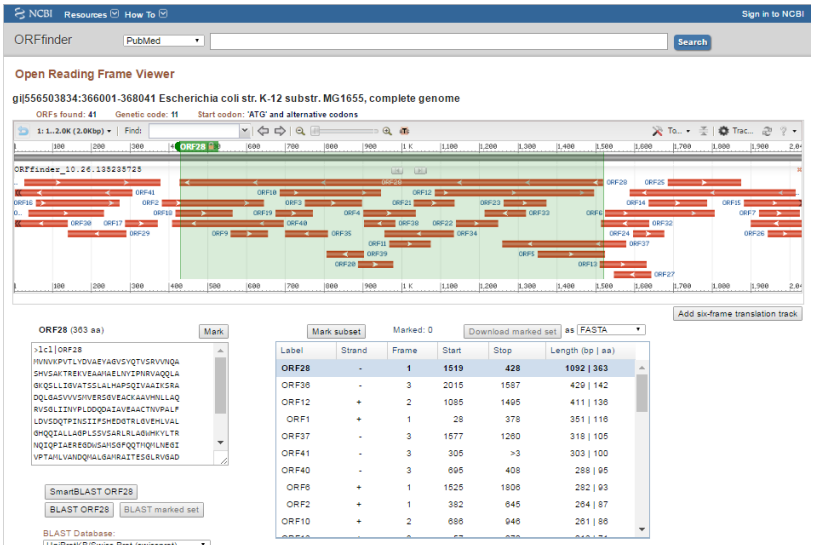
5.3. Escolher o código genético: ***11. Bacterial, Archaeal and Plant Plastid***



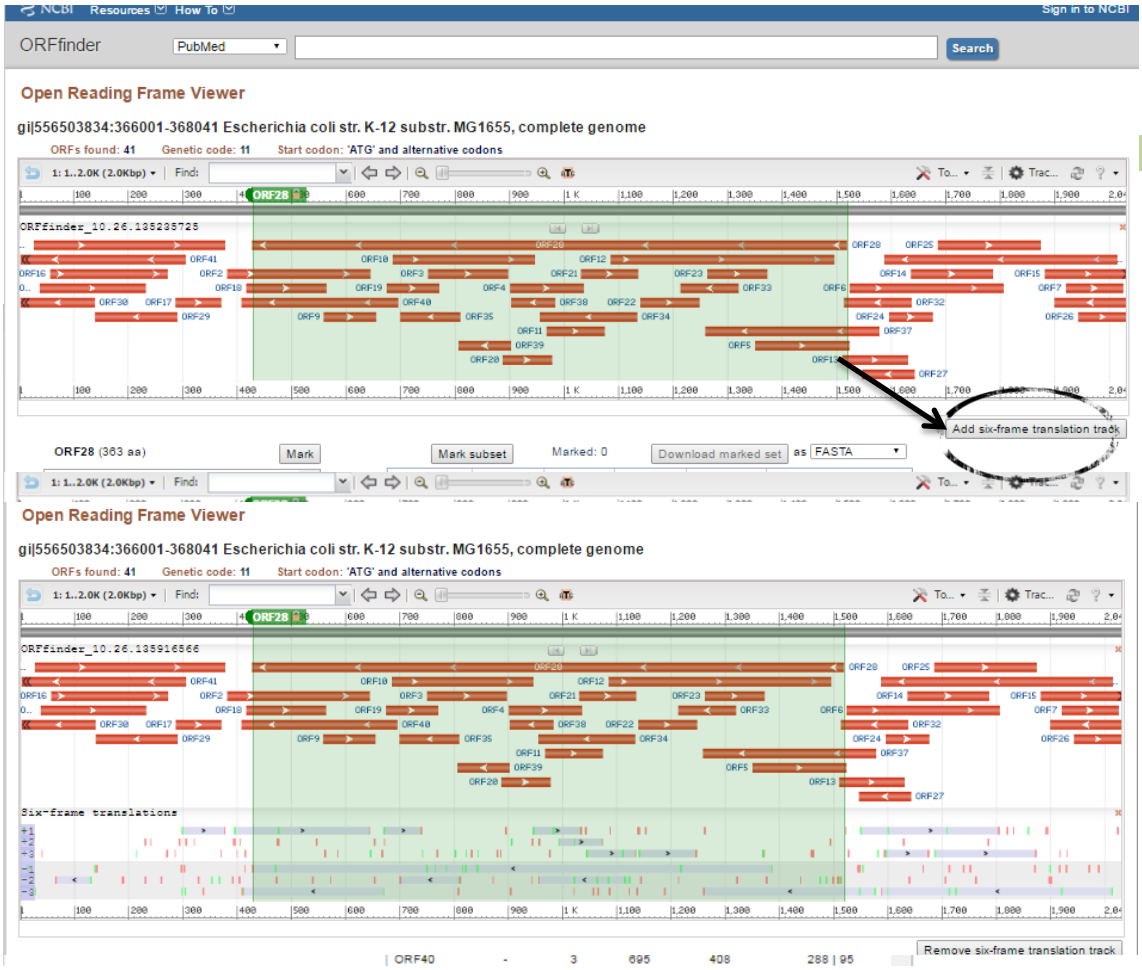
5.4. Escolher a opção: ***"ATG" and alternative initiation codons*** ("ATG" e codões de iniciação alternativos) e clicar em ***Submit*** (Submeter)



5.5. Analisar os resultados obtidos



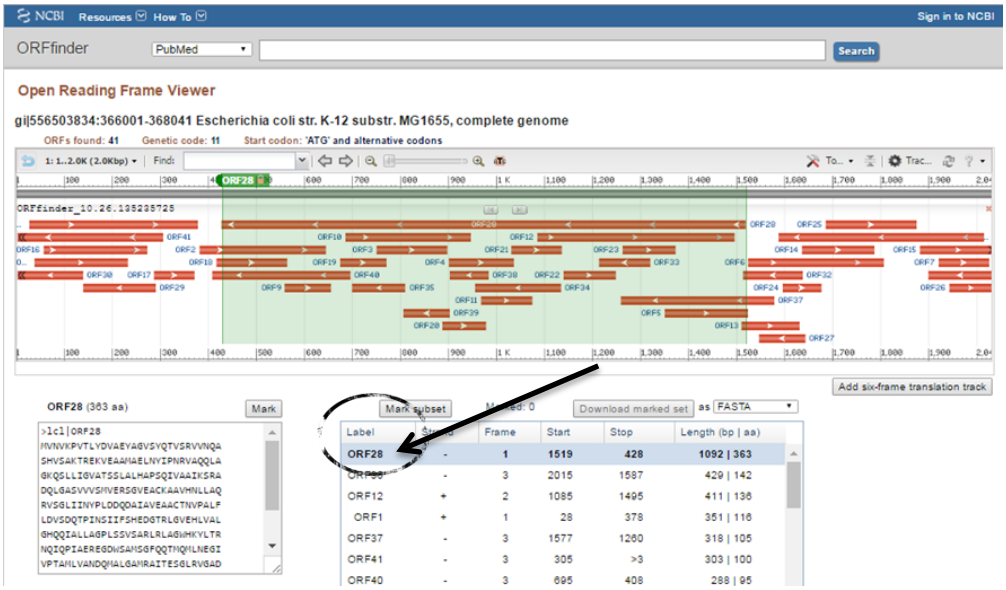
5.6. Abrir a opção ***Add six-frame translation track*** e explorar a função ***ATG*** (não fechar a página e prosseguir para o exercício 6)



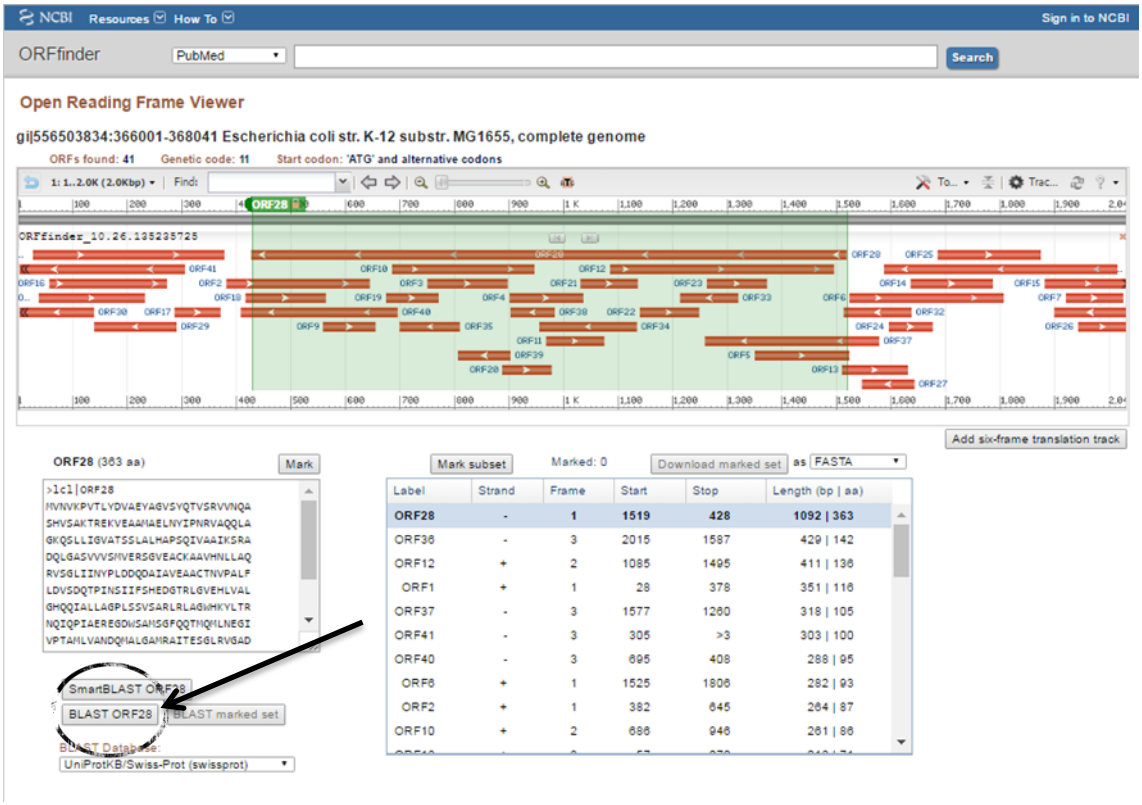
Exercício 6: *Das ORF's a genes putativos*

6.1. Analisar as características das ORF's obtidas.

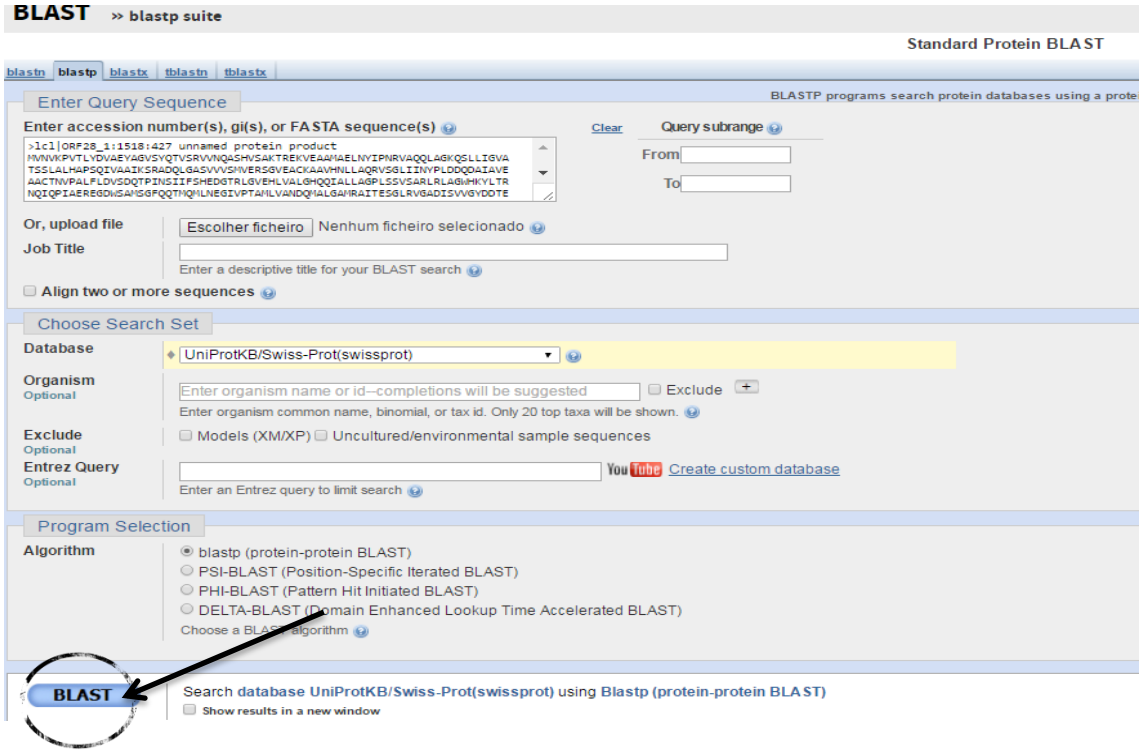
6.2. Selecionar uma ORF clicando no repetitivo código (exemplo: ORF28)



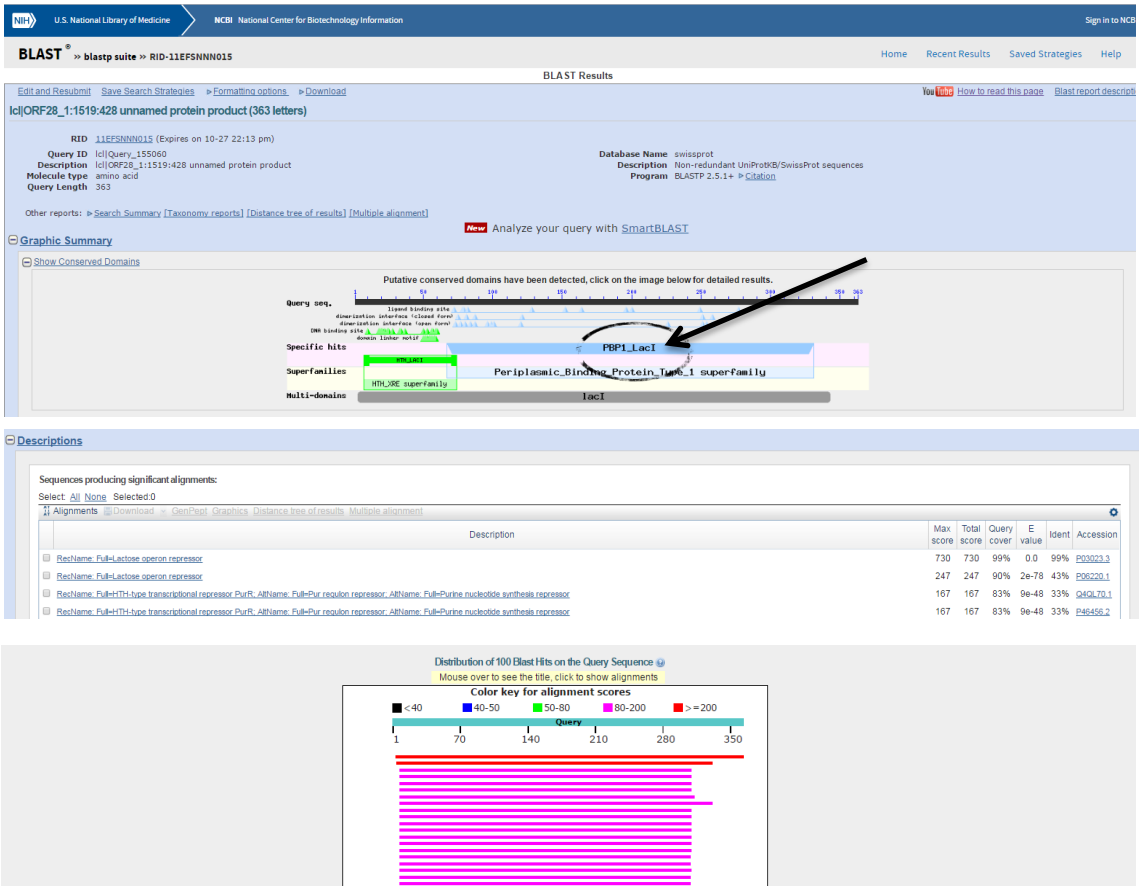
6.3. Iniciar o BLAST da ORF selecionada clicando em ***BLAST ORF*** (será redirecionado para uma página do *NCBI BLAST tool*: https://blast.ncbi.nlm.nih.gov/Blast.cgi)



6.4. Na página aberta clicar em ***BLAST*** para iniciar



6.5. Identificar o gene.



6.6. Repetir os passos 6.2. a 6.5. para duas ORF's alternativas.

Operão *lac*: Regulação Génica e Relações Evolutivas (Parte I)

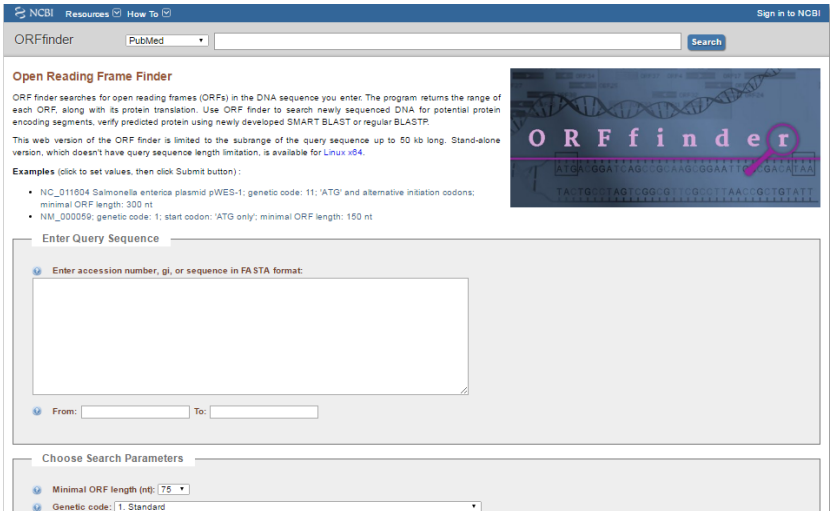
Com auxílio da aplicação de análise de ORF's NCBI ORF finder (http://www.ncbi.nlm.nih.gov/orffinder/) é objetivo da primeira parte desta sessão a identificação de codões de iniciação. Analisando o código genético é possível encontrar diferentes codões de iniciação.

Enquadramento curricular:

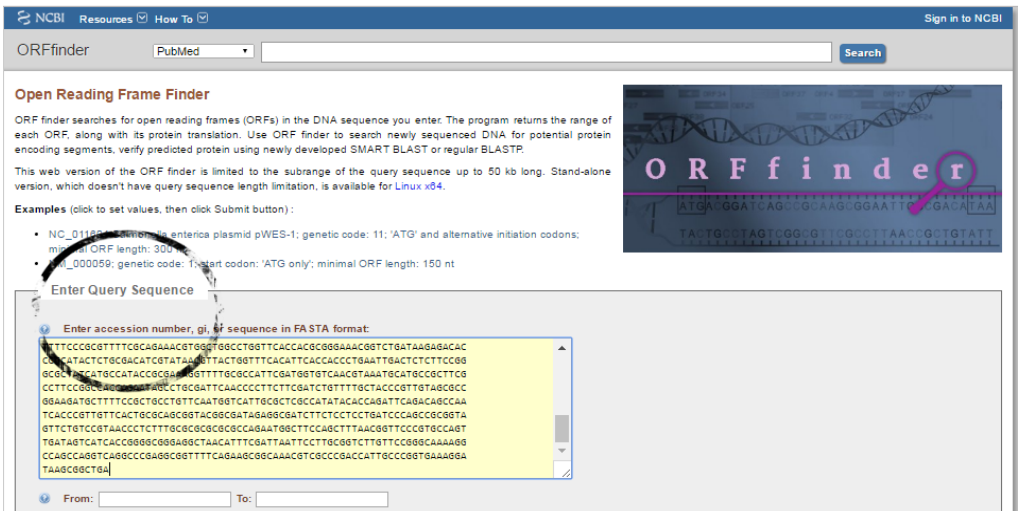
Será possível abordar temas enquadrados no 12ºano de escolaridade como a expressão genética, podendo também ser dinamizados no 11ºano de escolaridade com o objetivo de explorar os processos de transcrição e de tradução (Unidade 5: Crescimento e renovação celular – 1.1. DNA e Síntese Proteica), interpretando qual o possível ou possíveis codões de iniciação de um determinado gene (lac I).

Exercício 7: *Identificação de ORF's (codões de iniciação ATG)*

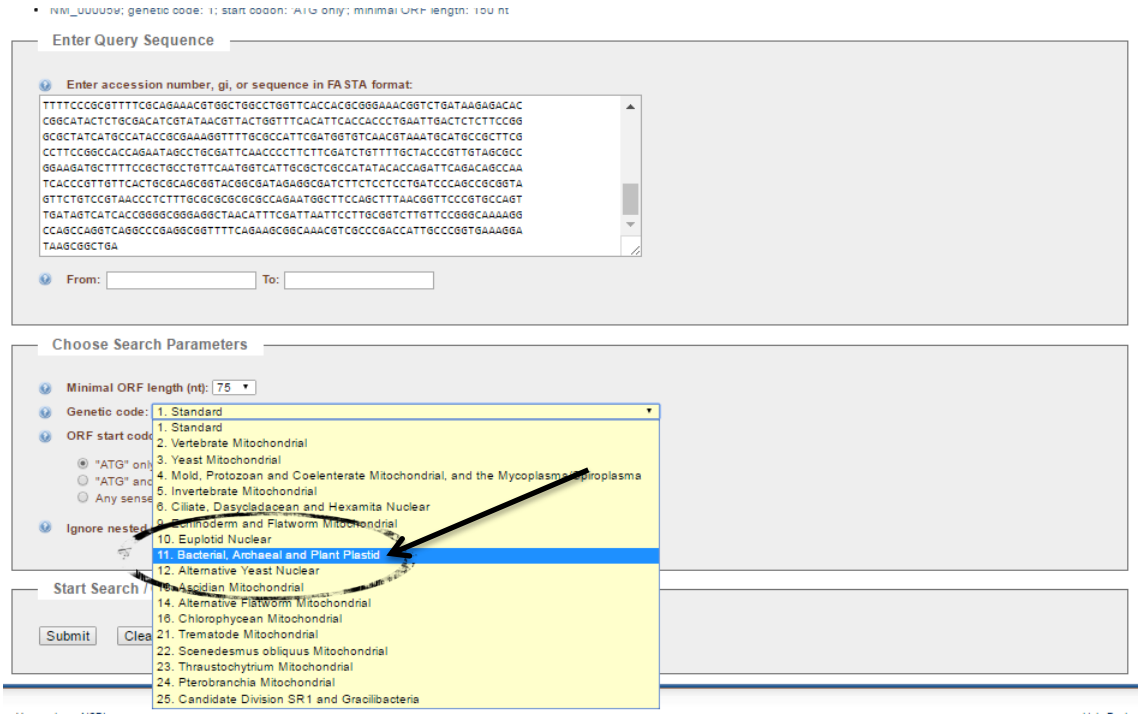
7.1. Aceder à página "*ORFfinder*" da plataforma NCBI através do link: http://www.ncbi.nlm.nih.gov/orffinder/



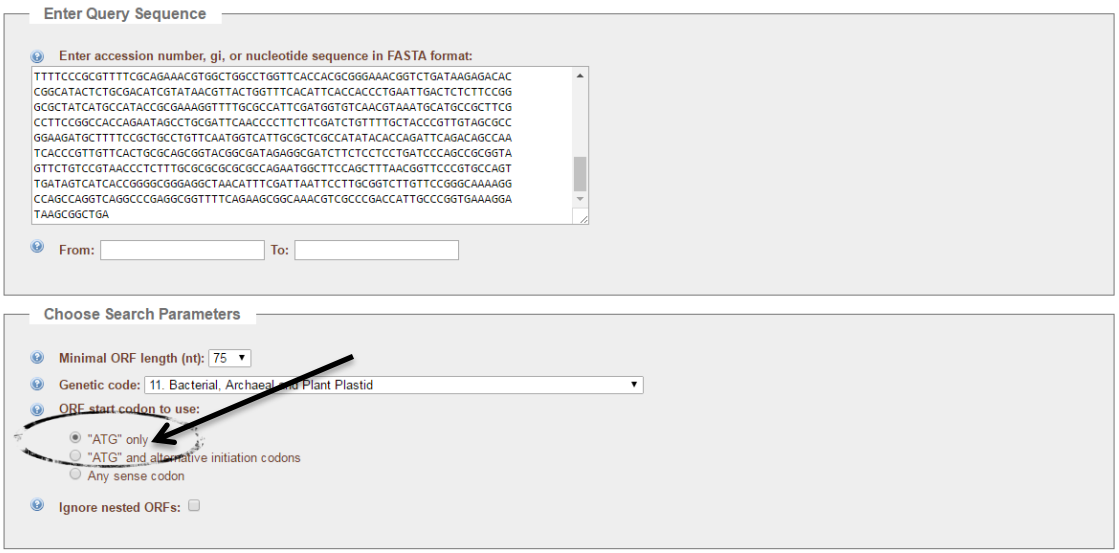
7.2. Inserir a sequência obtida previamente e anotada no processador de texto



7.3. Escolher o código genético: ***11. Bacterial, Archaeal and Plant Plastid***



7.4. Escolher a opção: ***"ATG" only*** e clicar em ***Submit*** (Submeter)



7.5. Analisar os resultados obtidos.

7.6. Repetir o Exercício 6 para ORF's geradas no exercício 7.4.

Operão *lac*: Regulação Génica e Relações Evolutivas (Parte II)

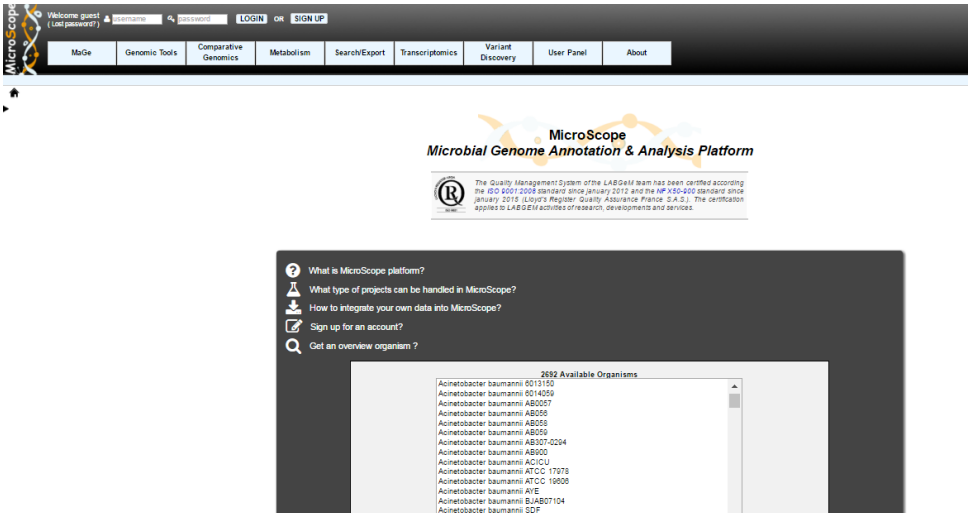
Utilizando uma plataforma bioinformática de genómica comparativa (MaGe – MicroScope - https://www.genoscope.cns.fr/agc/microscope/home/index.php) ser possível investigar relações evolutivas, identificando a presença dos genes do operão lac e das regiões flanqueantes em diferentes grupos taxonómicos. Este exercício permitirá elaborar hipóteses evolutivas que possam explicar a presença dos genes em taxa distintos.

Enquadramento curricular:

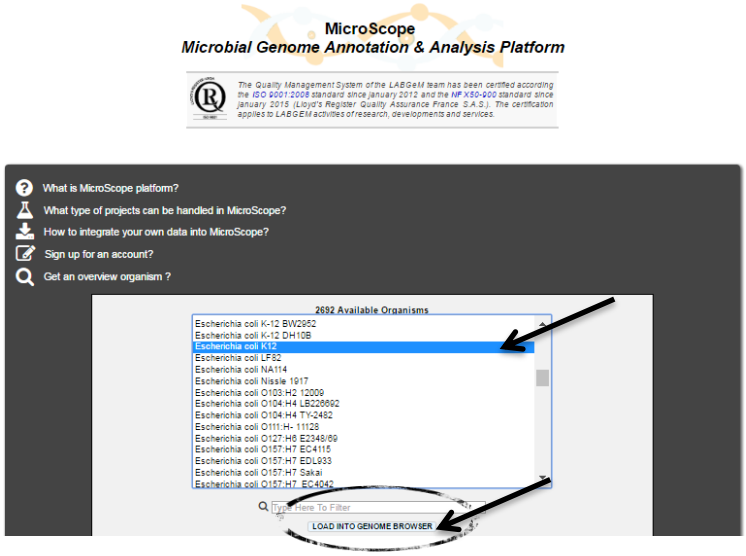
Será possível abordar no 11ºano de escolaridade a Unidade 7: Evolução Biológica.

Exercício 8: *Genómica Comparativa*

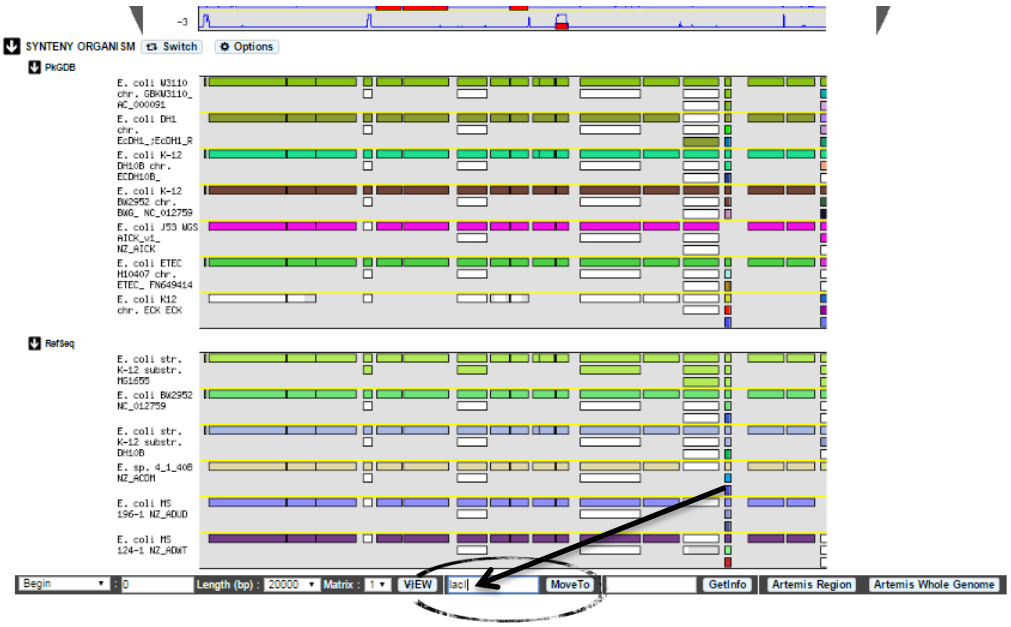
8.1. Aceder à plataforma MicroScope através do link: https://www.genoscope.cns.fr/agc/microscope/home/index.php



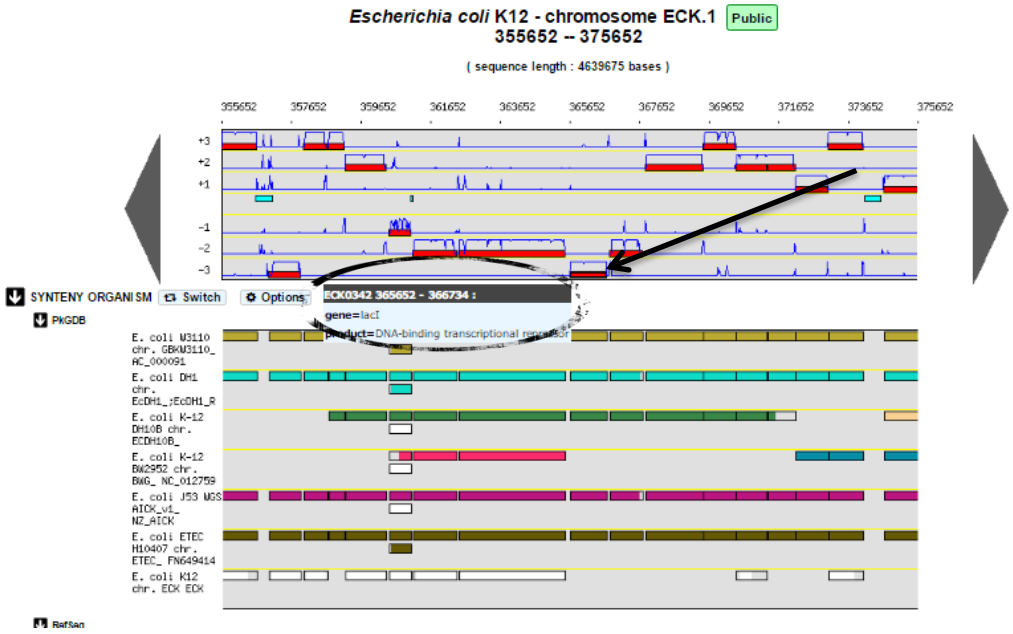
8.2. Escolher: ***Escherichia coli K12*** e clicar em ***load into genome browser*** (carregar no motor de busca de genomas)



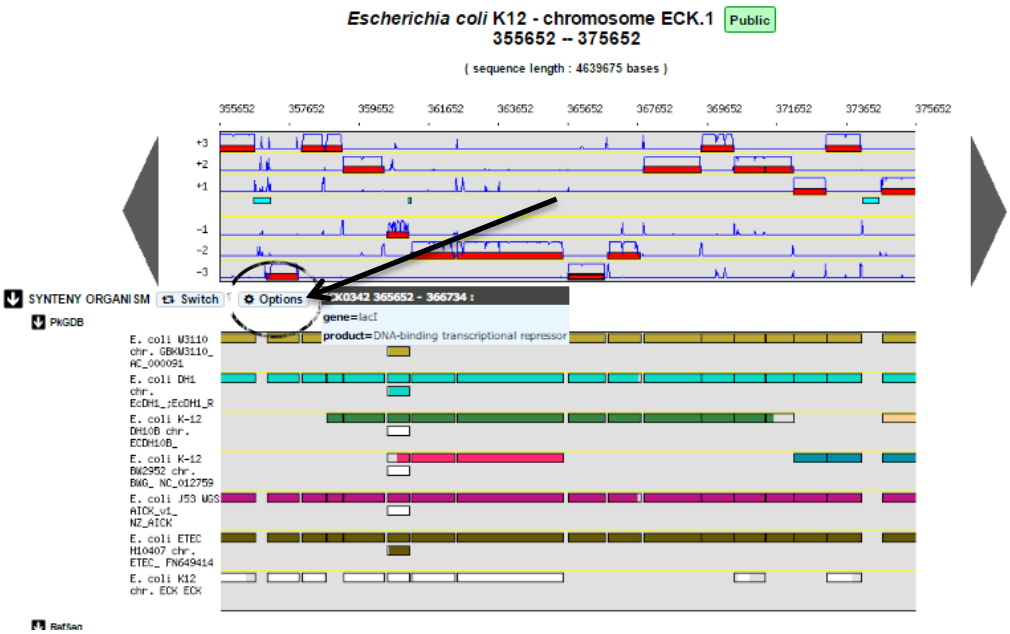
8.3. Para localizar o gene, pesquisar por ***lacI*** e clicar em: ***Move to*** (Mover para) na barra abaixo ilustrada.



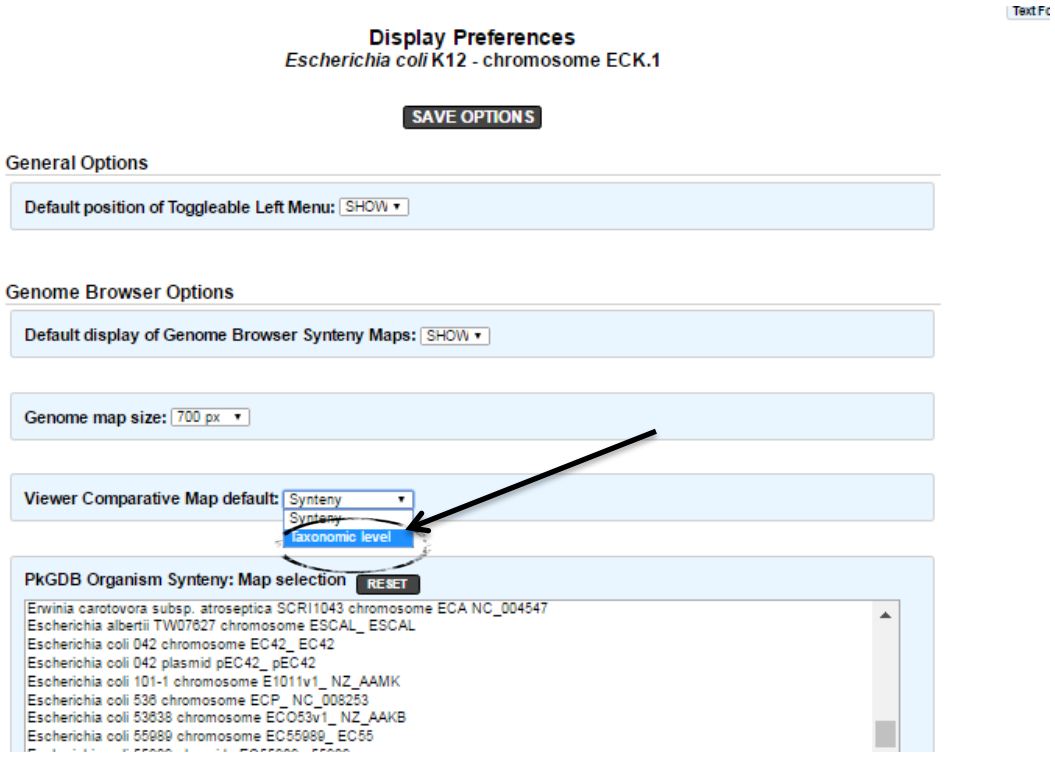
8.4. Identificar o gene *lacI* movendo o rato por cima das barras vermelhas



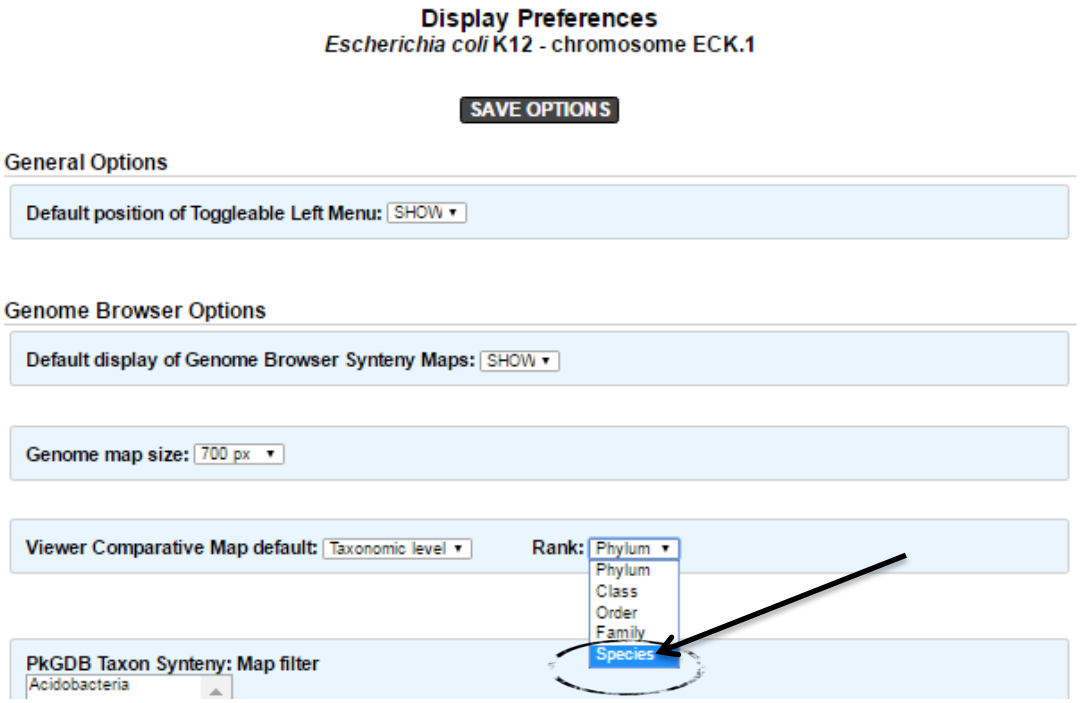
8.5. Selecionar ***Options Menu*** (Menu de opções)



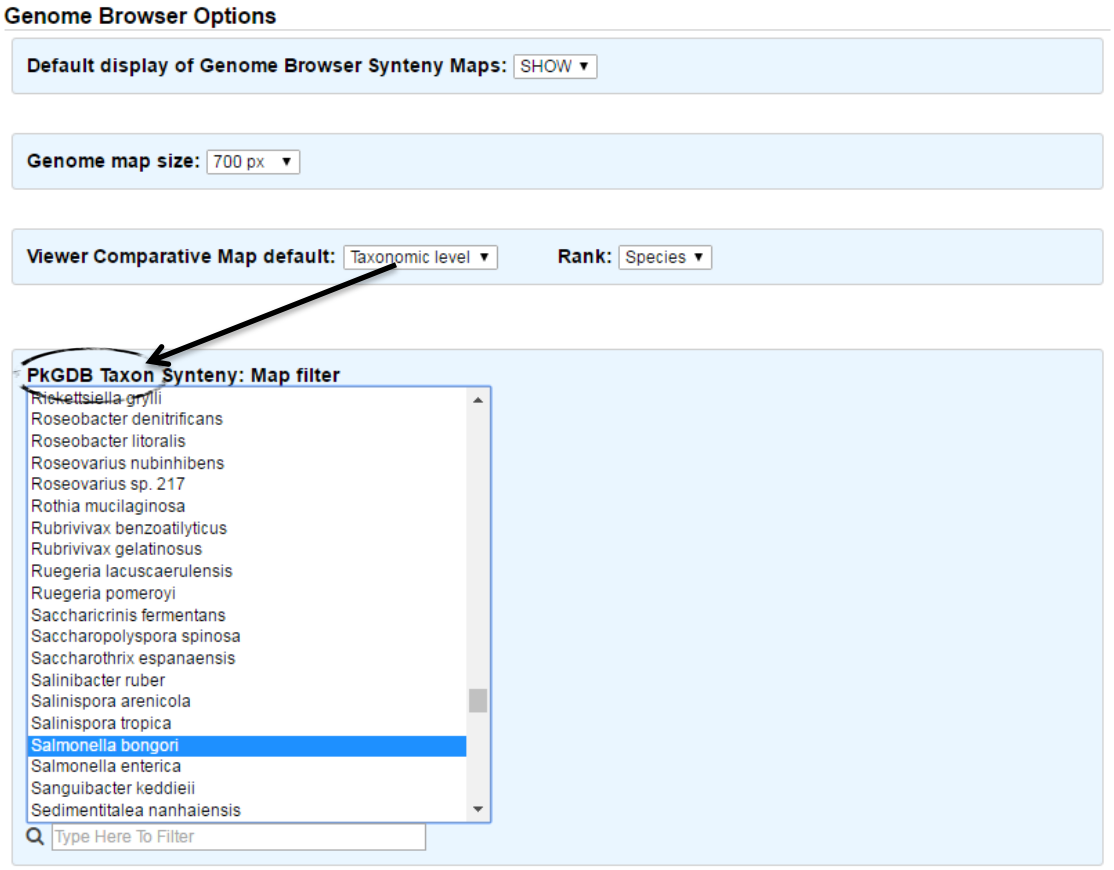
8.6. Na nova janela aberta, na secção ***Viewer Comparative Map default***, escolher ***taxonomy level*** (nível taxonómico)



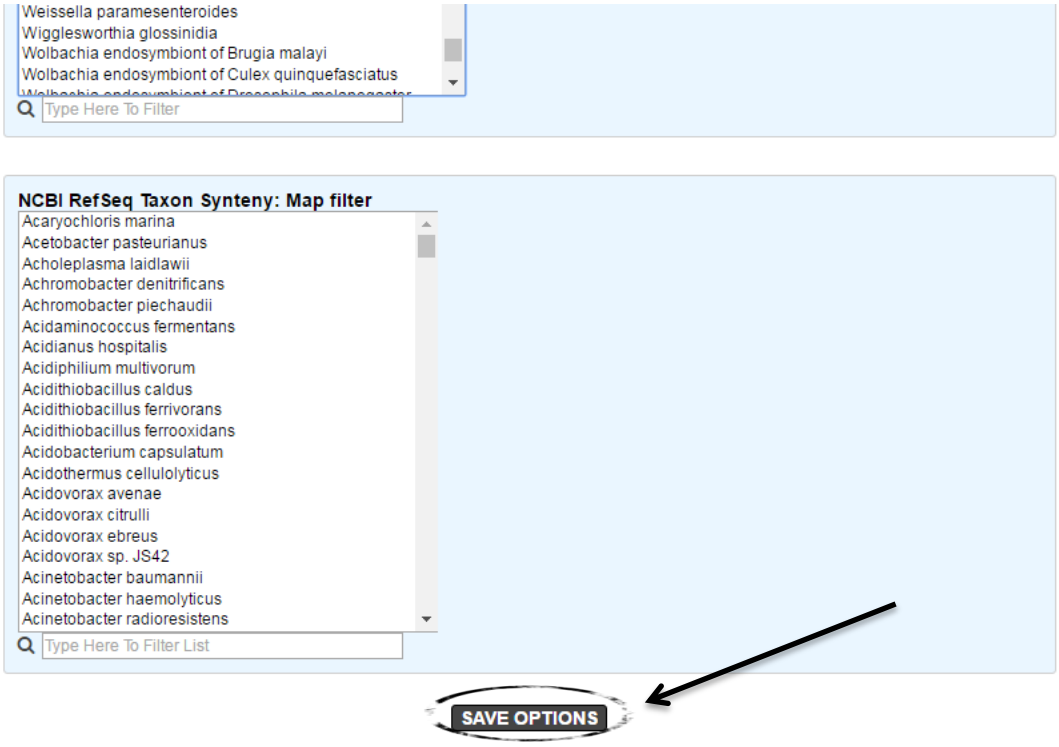
8.7. Em ***Rank*** (Ordenar), escolher ***Species*** (Espécies)



8.8. Na secção ***PkGDB Taxon Synteny: Map filter***, pressionar o botão CTRL e escolher *Bacillus cereus, Escherichia sp.4_1_40B, Samonella bongori e Vibrio fischeri.*



8.9. Clicar em ***Save options*** (Guardar opções)



8.10. Comparar a presença e a função do gene em diferentes *taxa*.

Importância dos Genes e das Mutações no Estudo da Evolução (Parte I)

Propõe-se a utilização de ferramentas bioinformáticas e bases de dados disponibilizados no EDGAR (https://edgar.computational.bio.uni-giessen.de/). Escolhendo até cinco estirpes bacterianas representantes dos grupos taxonómicos analisados na Parte II da sessão 3, pretende-se identificar o conjunto de genes homólogos e genes específicos de cada estirpe utilizando a funcionalidade "Diagramas de Venn" na plataforma EDGAR. Com base nos resultados obtidos discutir as noções de genoma centro (core genoma), pan genoma e genoma acessório. O EDGAR permitirá ainda identificar a sequência nucleotídica de cada gene e inferir a sequência aminoacídica, tendo em consideração a frequência de utilização de codões para diferentes organismos (consultando *codon usage table*).

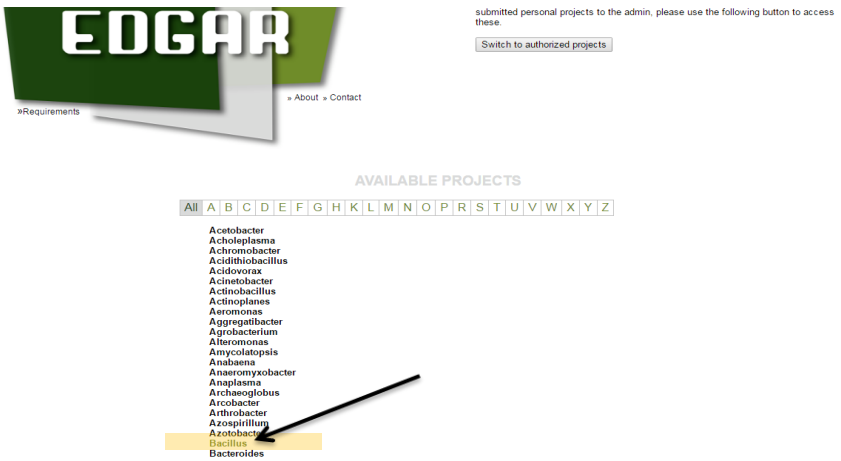
Enquadramento curricular:

Esta proposta enquadra-se particularmente no 11ºano de escolaridade na Unidade 5: Crescimento e renovação celular – 1.1. DNA e Síntese Proteica; Unidade 8: Sistemática dos Seres Vivos - 1.2. Taxonomia e Nomenclatura.

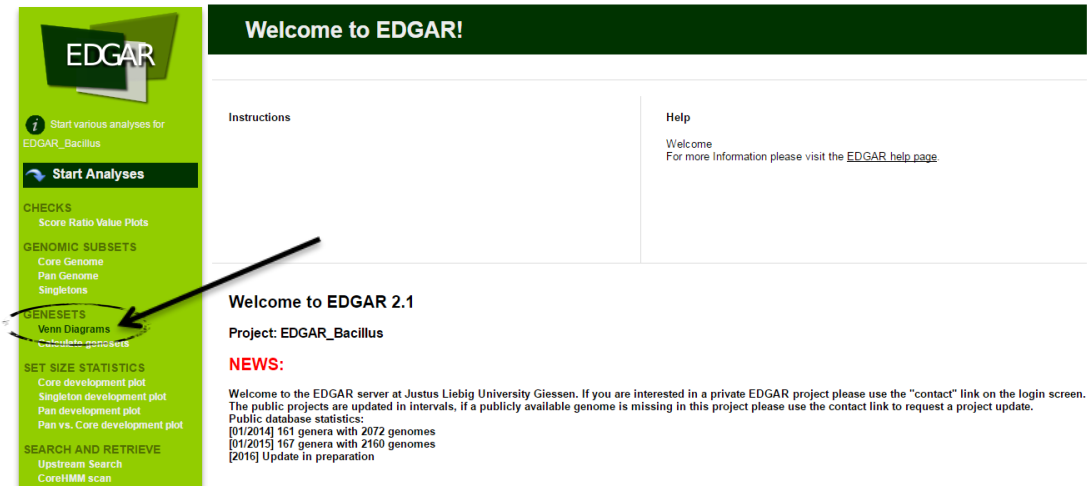
Exercício 9: *Explorando o EDGAR – Diagramas de Venn*

9.1. Aceder ao site: *https://edgar.computational.bio.uni-giessen.de/*

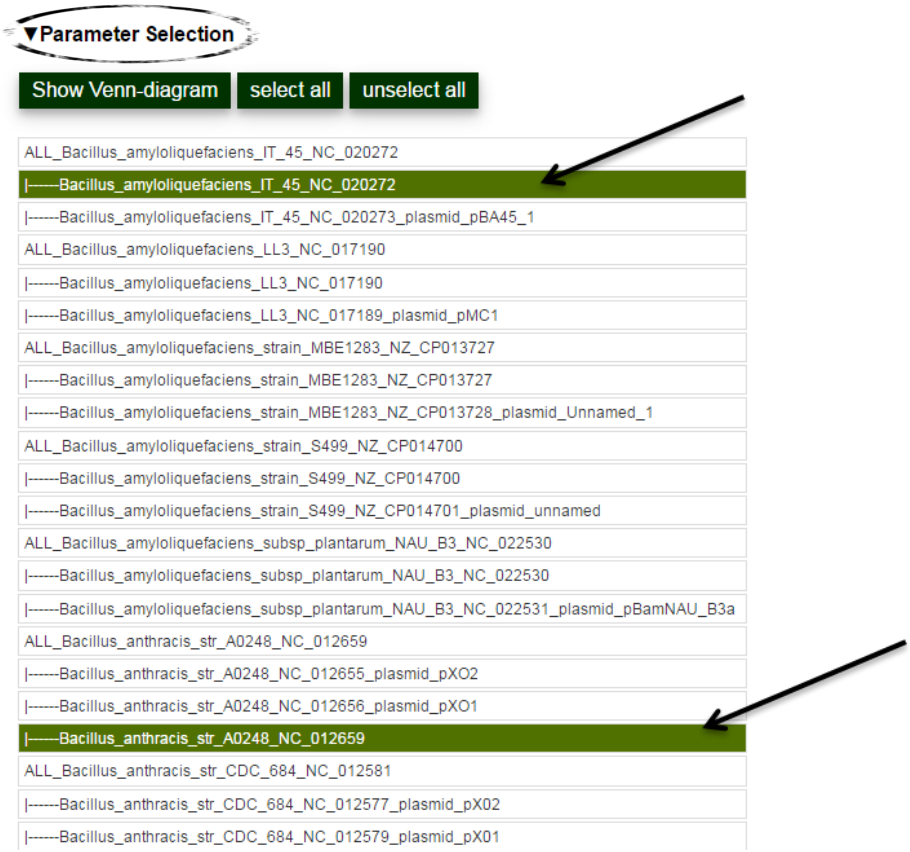
9.2. Escolher o género *Bacillus*



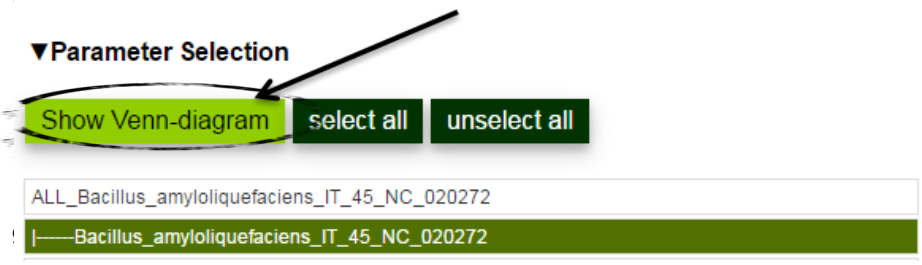
9.3. Selecionar: ***Venn Diagrams*** (diagrama de Venn).



9.4. Na secção ***Parameter Selection*** escolher as espécies ilustradas na imagem:

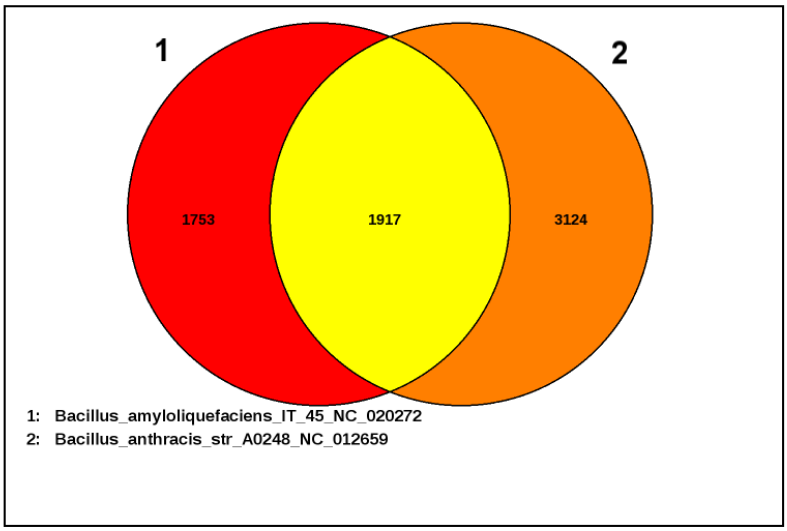


9.5. Selecionar: ***Show Venn-Diagram*** (Mostrar diagrama de Venn)



▼Parameter Selection

▼Results

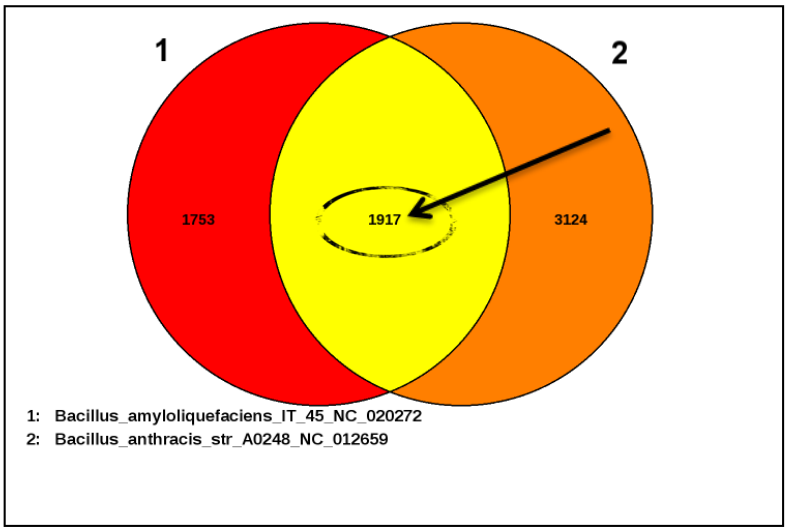


Gene list (names)

9.7. Selecionar os genes em comum.

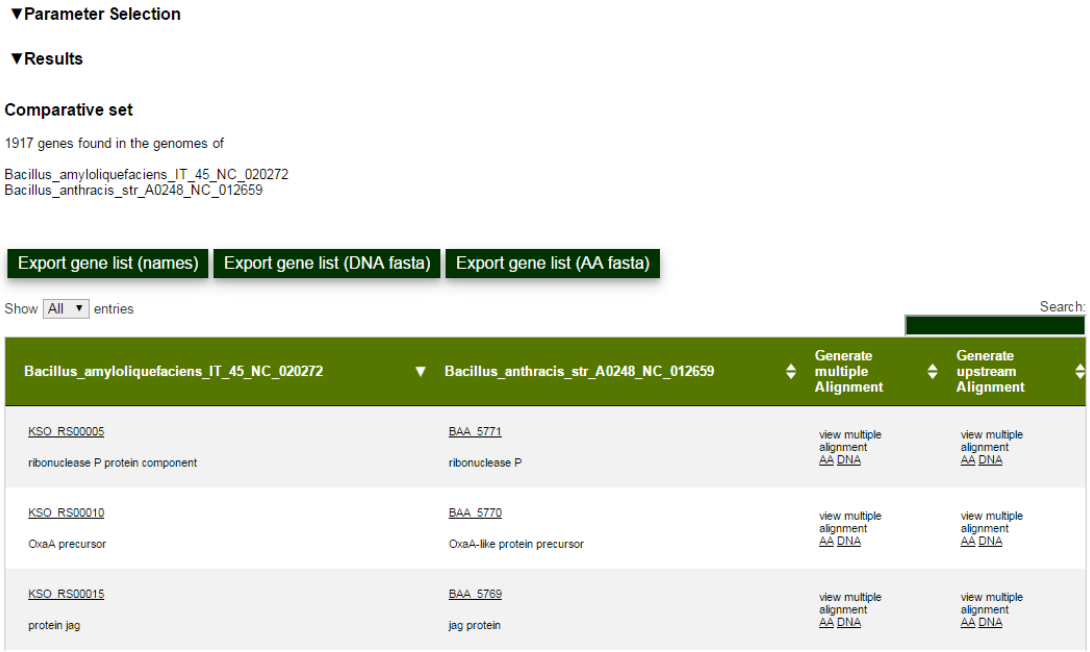
▼Parameter Selection

▼Results

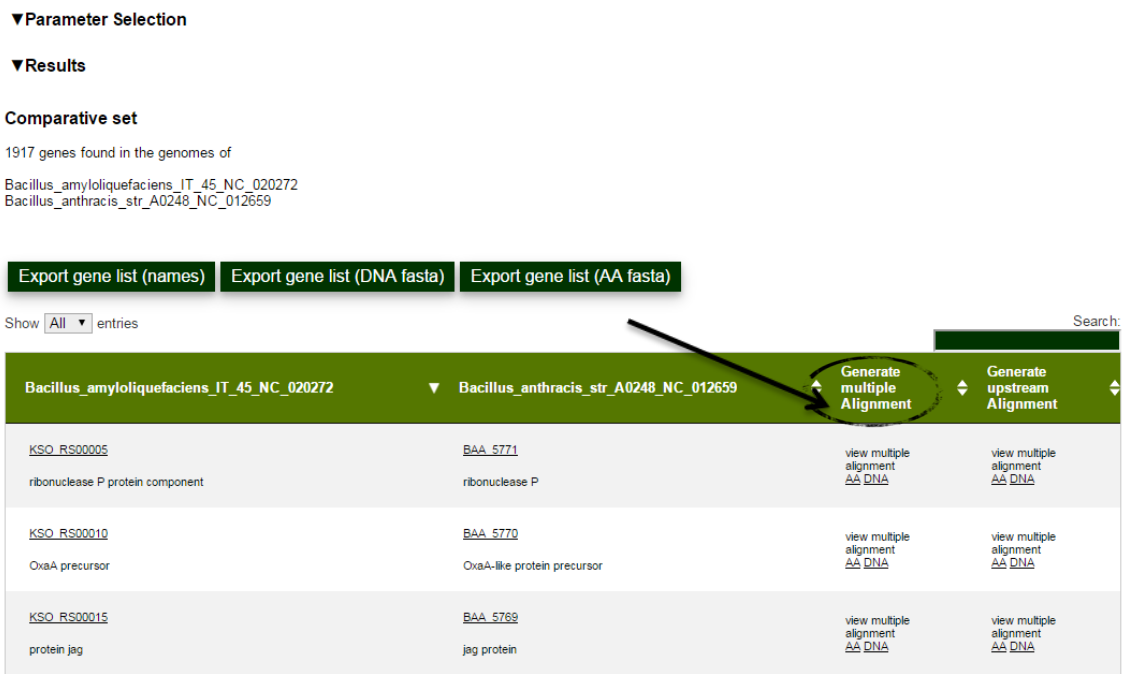


Gene list (names)

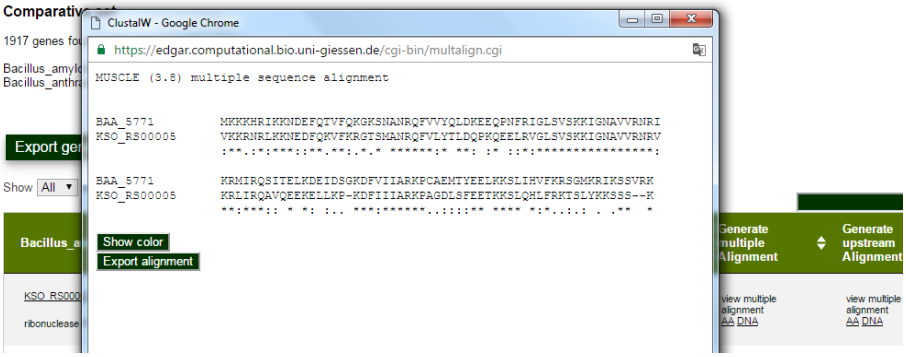
9.8. Analisar as semelhanças e diferenças de função entre os genes em comum.



9.9. Selecionar a opção ***aminoácidos (AA)*** na secção: ***Generate multiple alignment*** (Gerar alinhamentos múltiplos).



9.10. Identificar os aminoácidos produzidos (consultar *codon usage table*).



Codon usage table:

11. The Bacterial, Archaeal and Plant Plastid Code (transl_table=11)

TTT F Phe	TCT S Ser	TAT Y Tyr	TGT C Cys
TTC F Phe	TCC S Ser	TAC Y Tyr	TGC C Cys
TTA L Leu	TCA S Ser	TAA * Ter	TGA * Ter
TTG L Leu i	TCG S Ser	TAG * Ter	TGG W Trp
CTT L Leu	CCT P Pro	CAT H His	CGT R Arg
CTC L Leu	CCC P Pro	CAC H His	CGC R Arg
CTA L Leu	CCA P Pro	CAA Q Gln	CGA R Arg
CTG L Leu i	CCG P Pro	CAG Q Gln	CGG R Arg
ATT I Ile i	ACT T Thr	AAT N Asn	AGT S Ser
ATC I Ile i	ACC T Thr	AAC N Asn	AGC S Ser
ATA I Ile i	ACA T Thr	AAA K Lys	AGA R Arg
ATG M Met i	ACG T Thr	AAG K Lys	AGG R Arg
GTT V Val	GCT A Ala	GAT D Asp	GGT G Gly
GTC V Val	GCC A Ala	GAC D Asp	GGC G Gly
GTA V Val	GCA A Ala	GAA E Glu	GGA G Gly
GTG V Val i	GCG A Ala	GAG E Glu	GGG G Gly

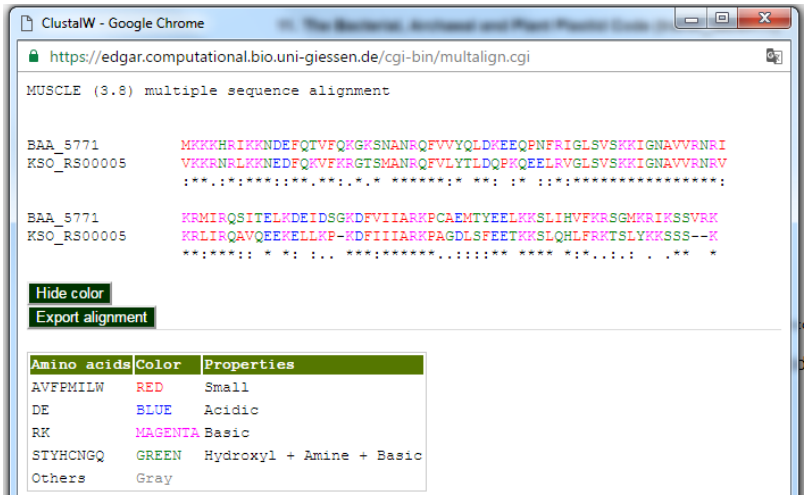
(retirado de: https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi?chapter=tgencodes#SG11)

O que significam os símbolos de consenso no alinhamento?

***** (asterisco) - indica posições que possuem um único resíduo, totalmente conservado.
: (dois pontos) - indica conservação entre grupos de propriedades fortemente semelhantes.
. (ponto) - indica conservação entre grupos de propriedades fracamente semelhantes.

*Nota – Os mesmos símbolos são exibidos para os alinhamentos DNA / RNA. No entanto deve ser tido em conta que apenas os caracteres * (asterisco) devem ser tidos em conta, sendo que os outros caracteres devem ser ignorados para os alinhamentos DNA / RNA.*

9.12. Clicar em ***Show Color*** (Mostrar cor) e explorar as características de cada aminoácido.

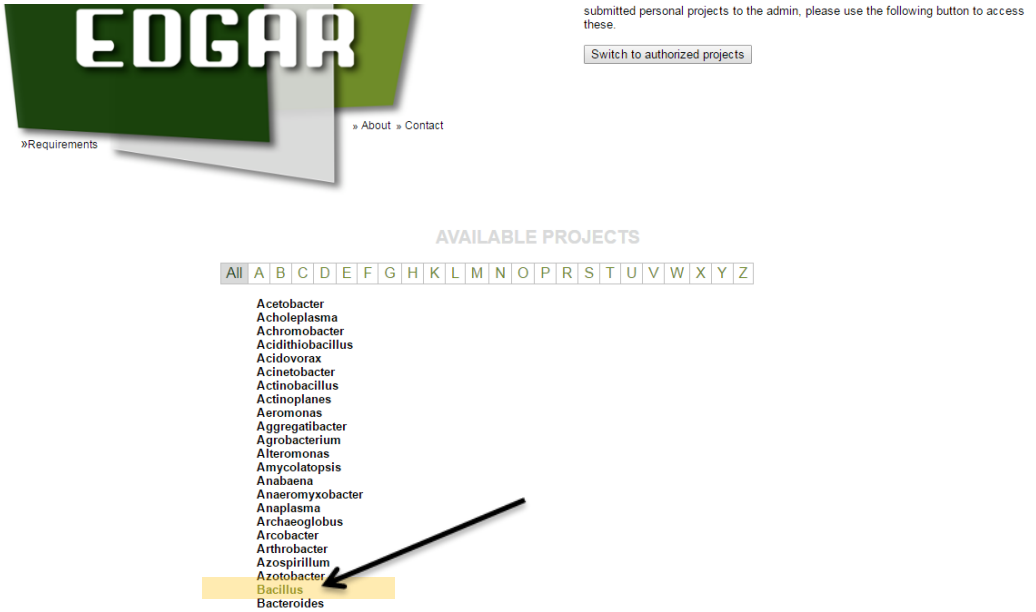


9.13. No ponto 9.4. selecionar 5 espécies e repetir os passos 9.5. a 9.8.

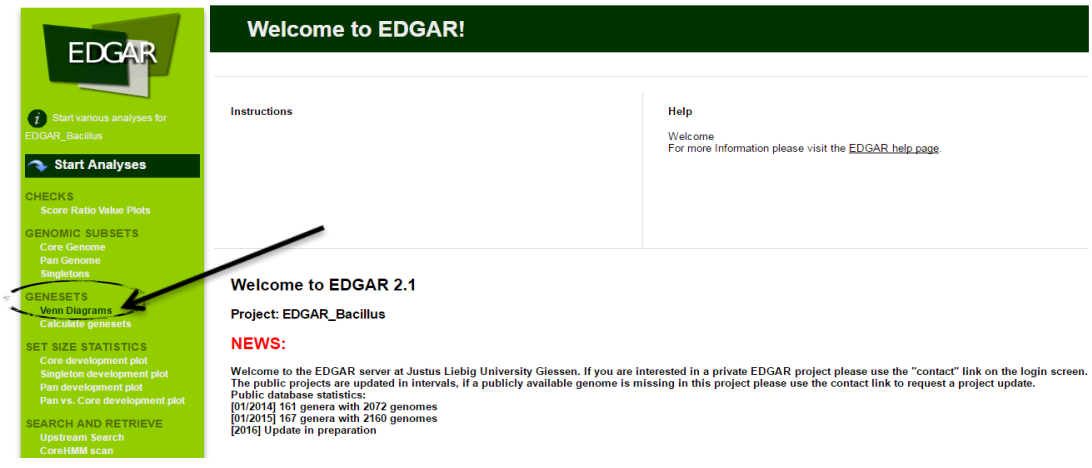
Exercício 10: *Explorando o EDGAR – Diagramas de Venn (Cromossomas vs. Plasmídeos)*

10.1. Aceder ao site: *https://edgar.computational.bio.uni-giessen.de/*

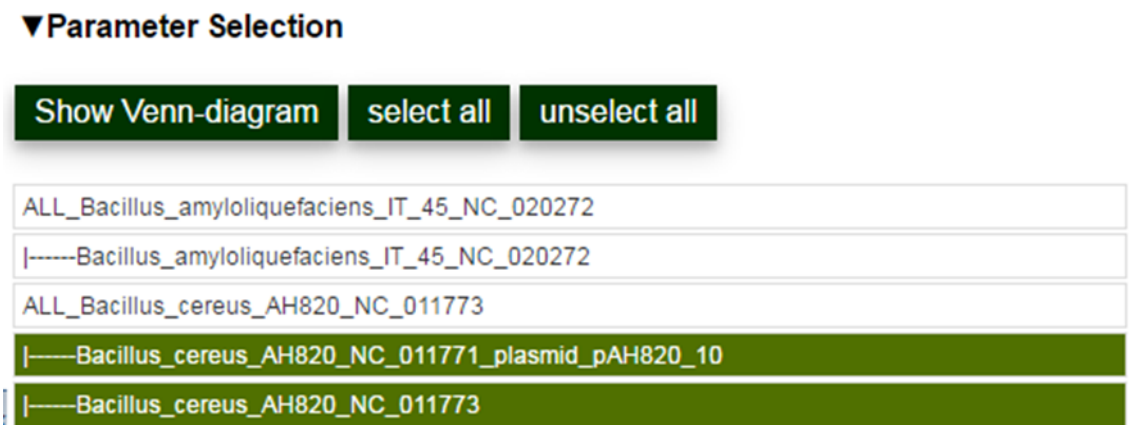
10.2. Escolher o género *Bacillus*



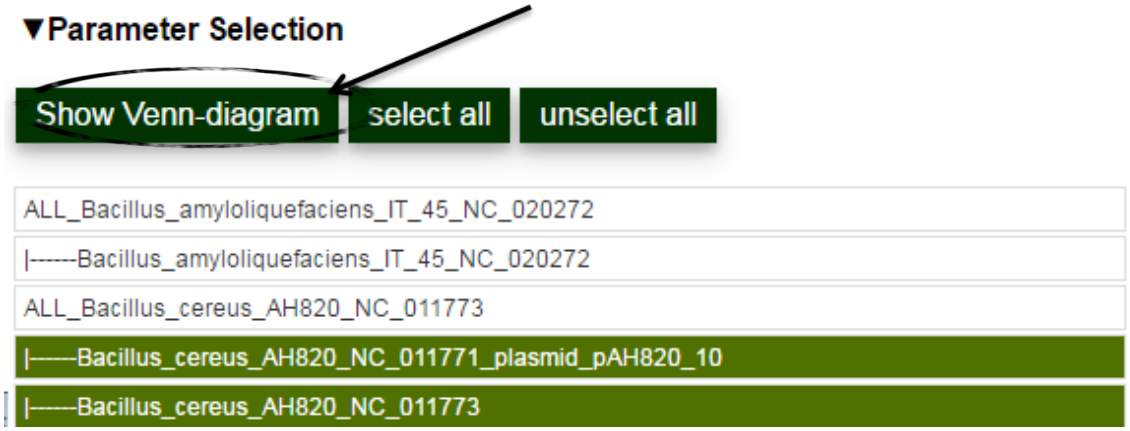
10.3. Selecionar: ***Venn Diagrams*** (Diagramas de Venn)**.**



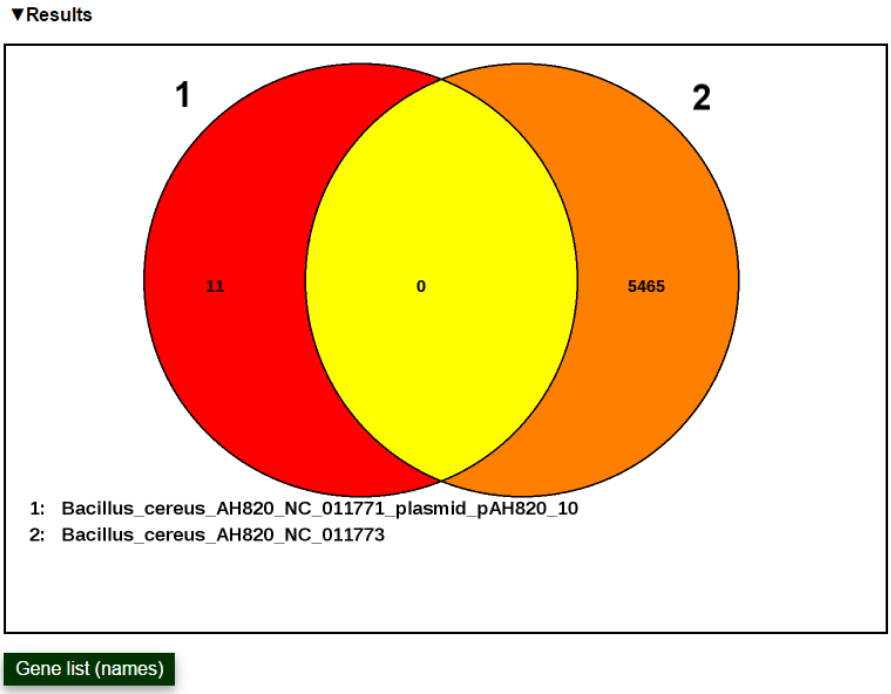
10.4. Na secção ***Comparasion sets*** escolher a espécie ilustrada na imagem:



10.4. Clicar em ***Show Venn-diagram*** (Mostrar Diagramas de Venn)



10.5. Concluir acerca da complementaridade do plasmídeo e do cromossoma principal.



Importância dos Genes e das Mutações no Estudo da Evolução (Parte II)

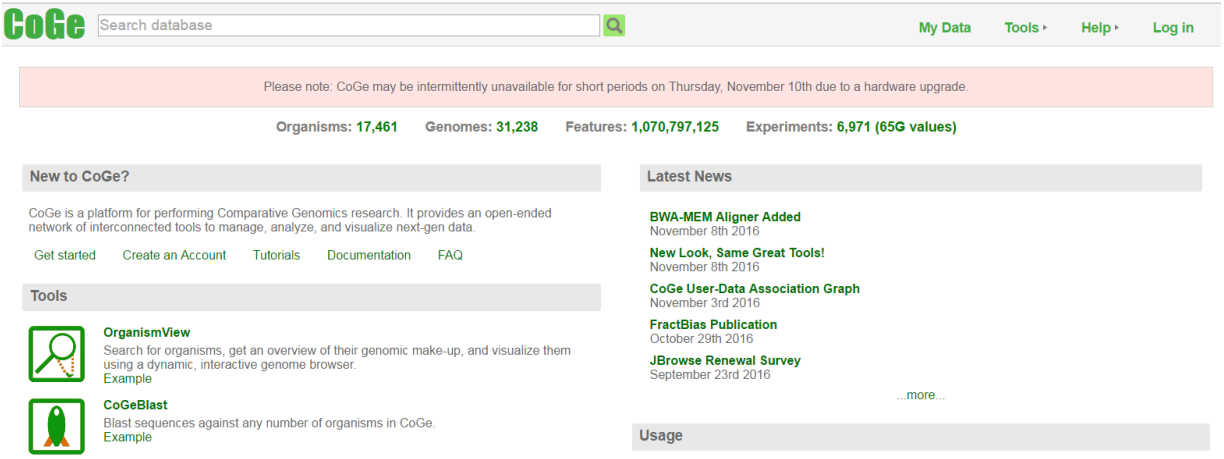
Estabelecer *plots* e mapas de sintenia através da comparação de genomas de duas das estirpes analisadas previamente, utilizando essencialmente o SynMap da plataforma CoGe (https://genomevolution.org/CoGe/). Neste exercício para além da noção de sintenia, os formandos poderão observar e registar importantes eventos de dinâmica genómica, nomeadamente inversões, deleções e inserções.

Enquadramento curricular:

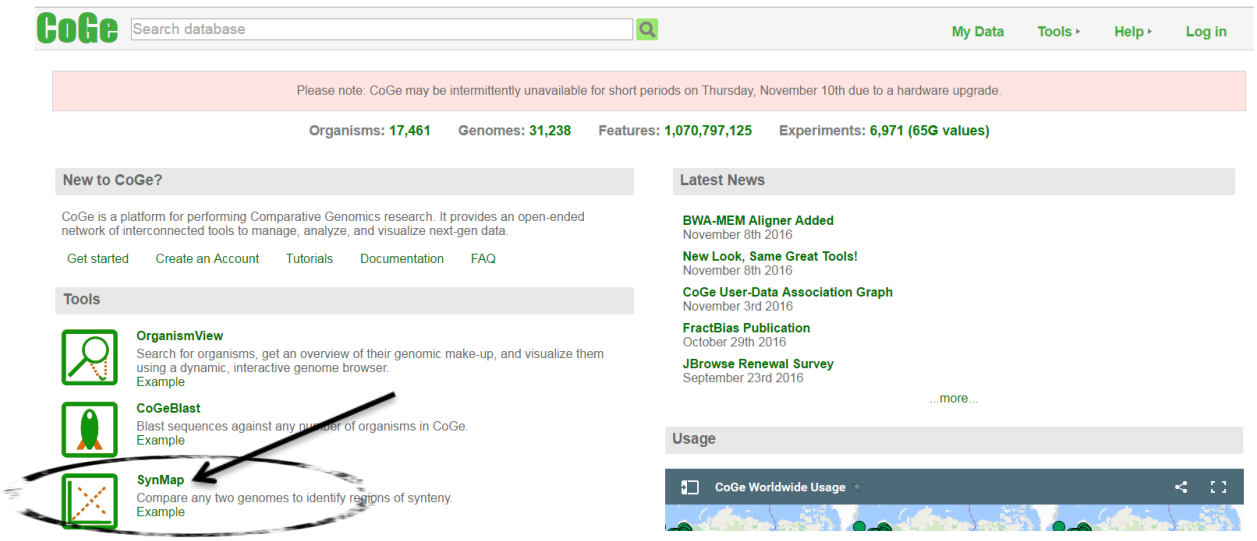
Permite-se a abordagem a conceitos de evolução enquadrados no 11ºano de escolaridade (Unidade 7 – 2. Mecanismos de Evolução), mas também relativos à compreensão da utilidade das mutações na exploração do tema 4. Alteração do material genético – 4.1. Mutações (Módulo 1) do programa 12ºano de escolaridade.

Exercício 11: *CoGe – Plots e Mapas de Sintenia*

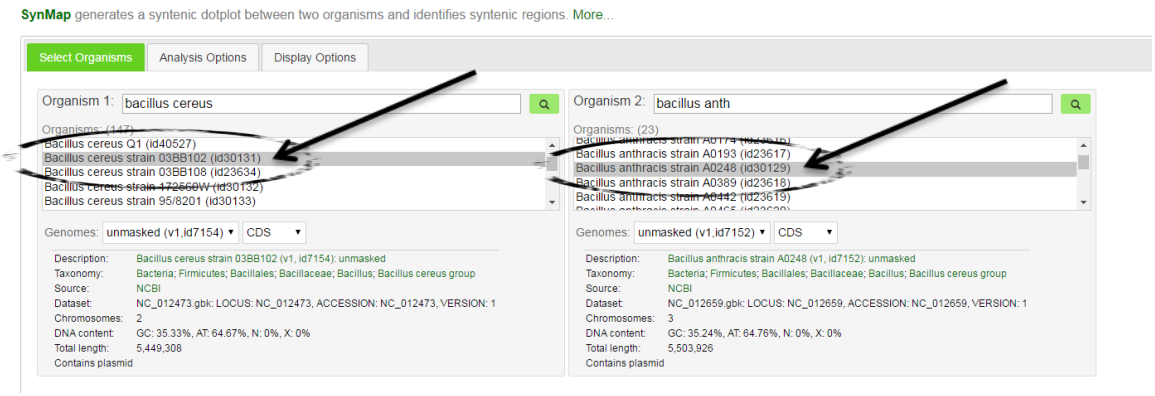
11.1. Aceder ao site: https://genomevolution.org/coge/



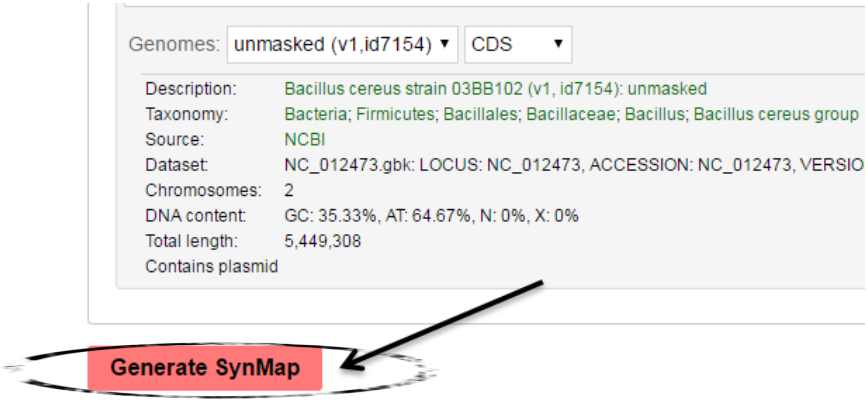
11.2. Selecionar a opção *SynMap*.



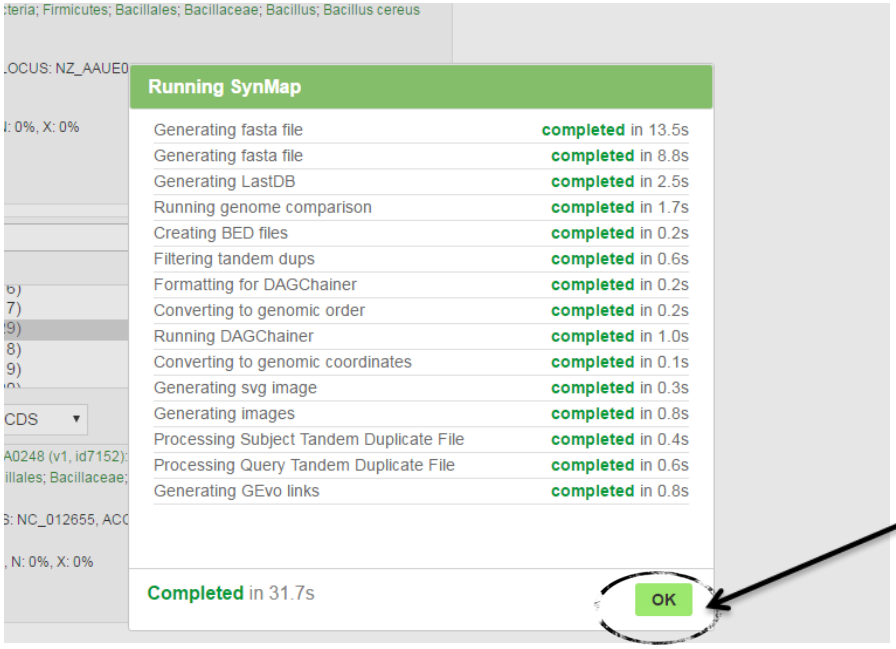
11.3. Selecionar como organismo 1: "*Bacillus cereus* 03BB102" e como organismo 2: "*Bacillus anthracis* A0248".



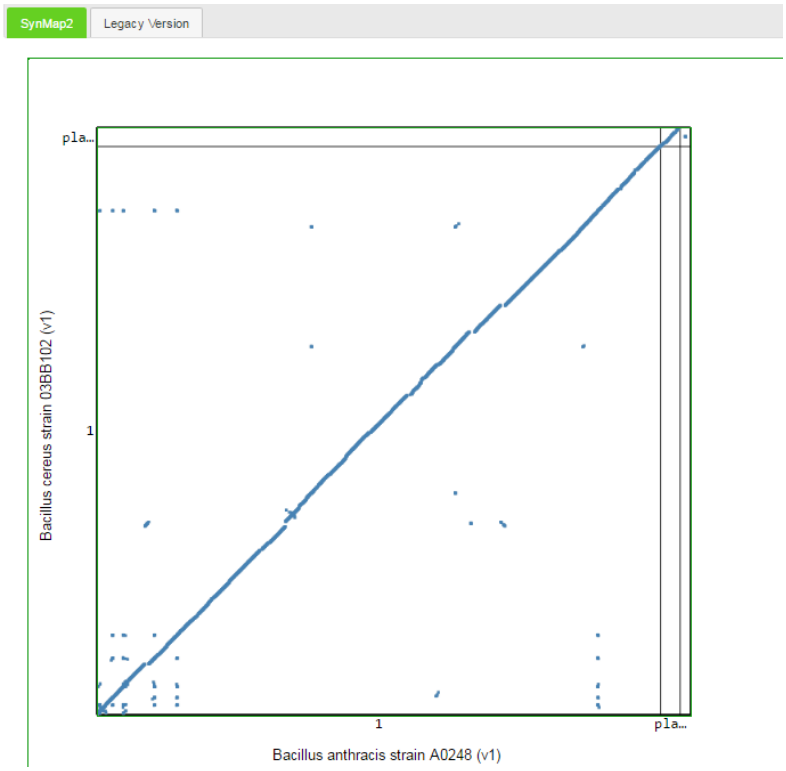
11.4. Clicar em: ***Generate SynMap*** (Gerar *SynMap*)



11.5. Na janela ***Running SynMap*** (Gerar SynMap) clicar em ***OK*** quando concluído.

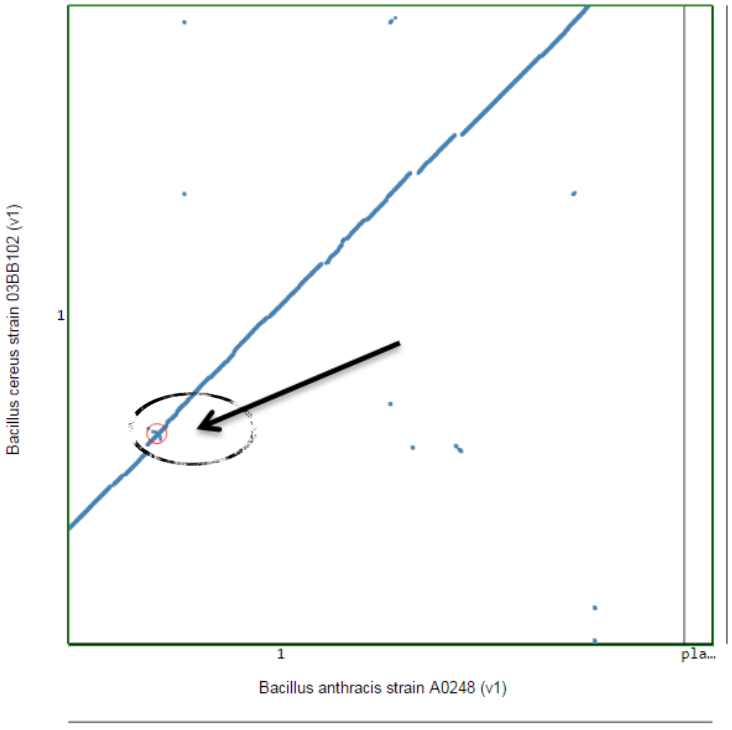


11.6. Identificar diferenças através da leitura do *SynMap*.

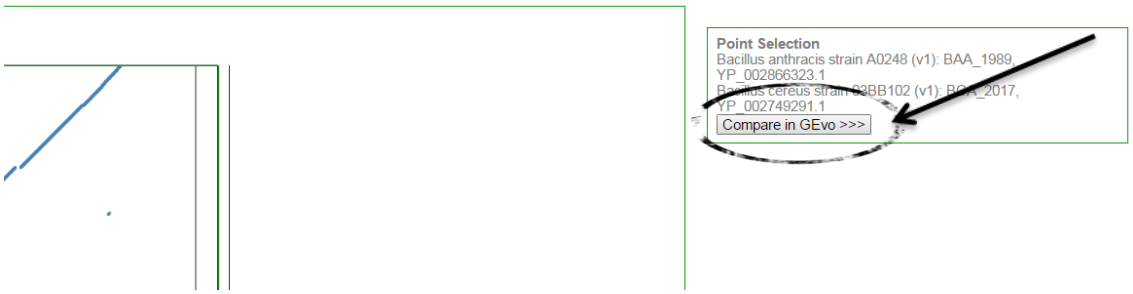


11.7. Comparar evolutivamente as espécies analisadas.

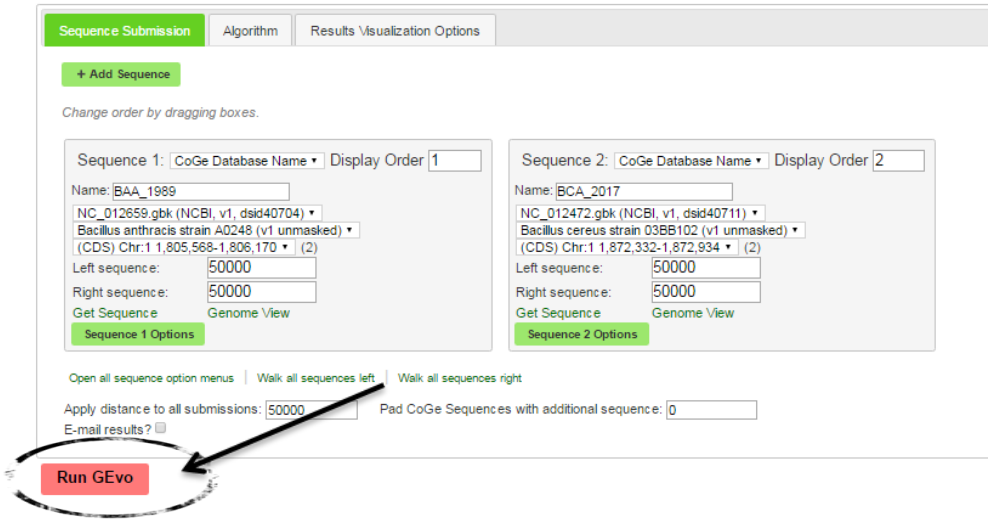
11.8. Selecionar um ponto do no mapa de sintenia.



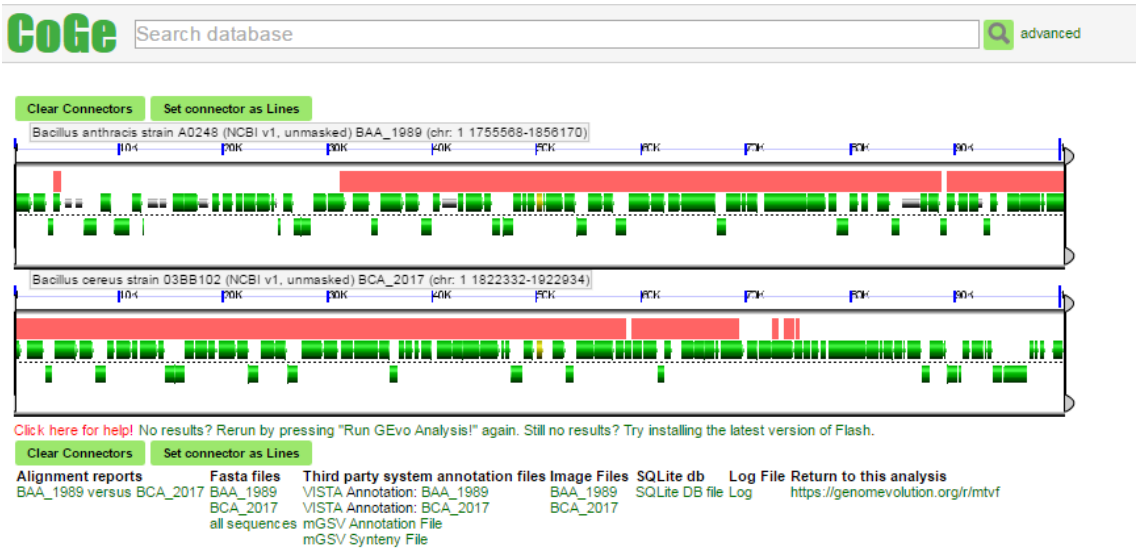
11.9. Selecionar a opção ***Compare in GEvo*** (Comparar em GEvo)



11.10. Clicar em ***Run GEvo*** (Iniciar GEvo)



11.11. Analisar as diferenças entre as duas espécies bacterianas para a zona selecionada.



Explorando Vias Metabólicas nos Diferentes Domínios da Vida (Parte I)

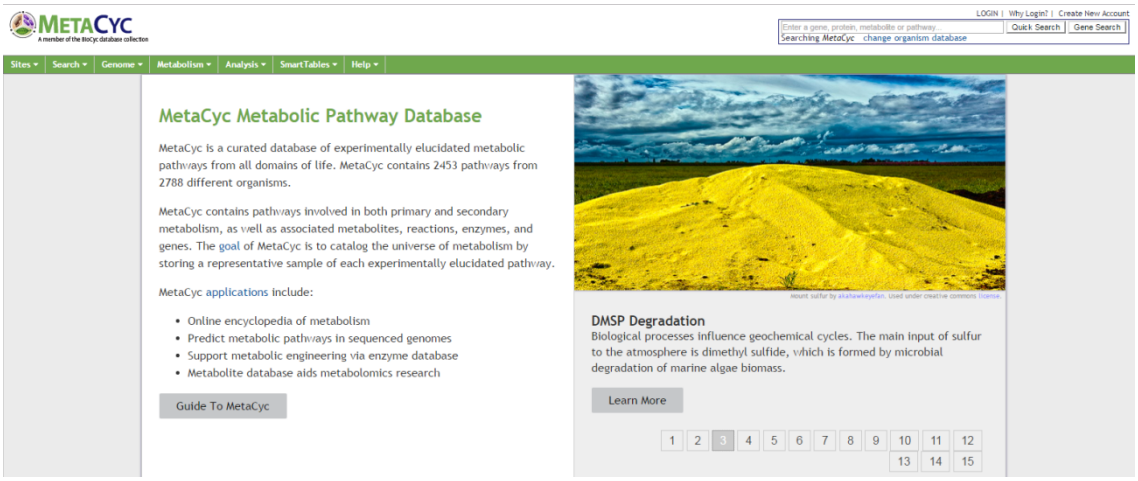
É objetivo a exploração da ferramenta de análise *MetaCyc - MetaCyc Metabolic Pathway Database* (http://metacyc.org/) que permitirá o estudo de reações do ciclo de Krebs (TCA cycle), e dos compostos envolvidos, com vista a identificar diferenças neste ciclo para diferentes domínios da vida, como os Eukarya e Bacteria. Este exercício será ainda um valioso contributo para compreender o carácter anfibólico dos compostos intermediários do ciclo de Krebs e o seu envolvimento noutras reações do metabolismo celular. Com este exercício os formandos poderão fundamentar a biossíntese de moléculas precursoras de outras biomoléculas essenciais como aminoácidos, purinas, pirimidinas e ácidos gordos.

Enquadramento curricular:

Esta atividade enquadra-se curricularmente no 10ºano de escolaridade: Unidade 3 – Transformação e utilização de energia pelos seres vivos: 1. Fermentação; 2. Respiração aeróbia.

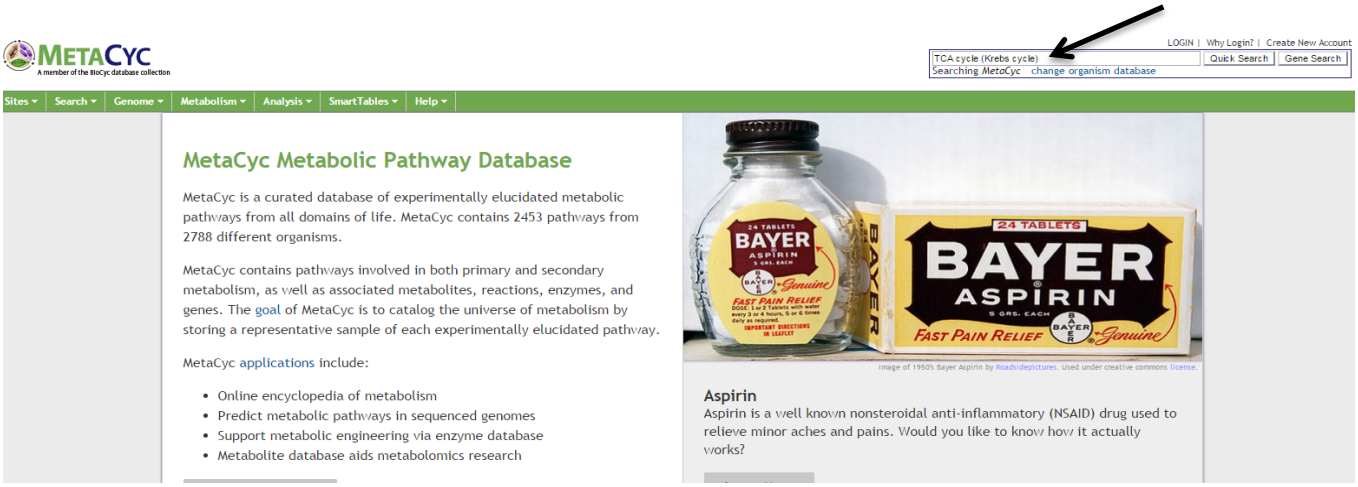
Exercício 12: *MetaCyc - MetaCyc Metabolic Pathway Database*

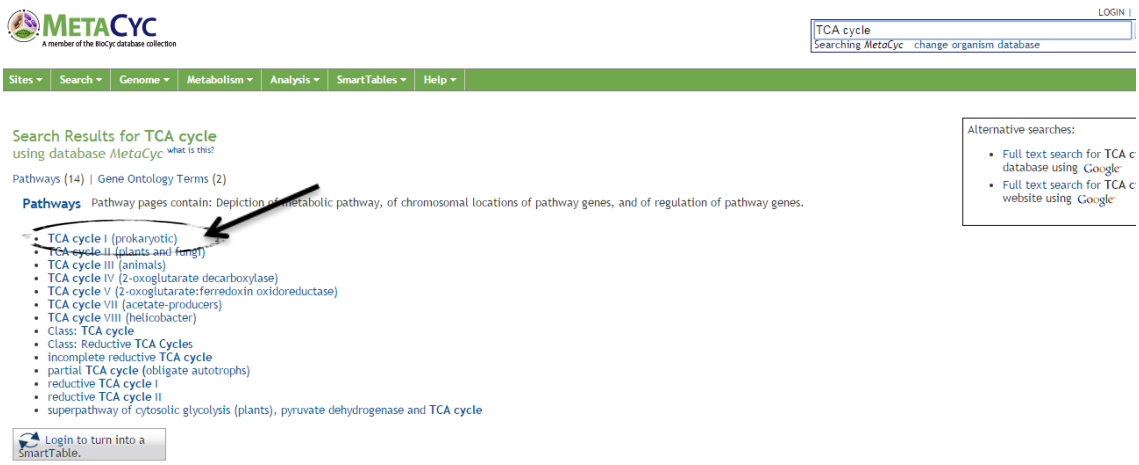
12.1. Aceder ao site: http://metacyc.org/



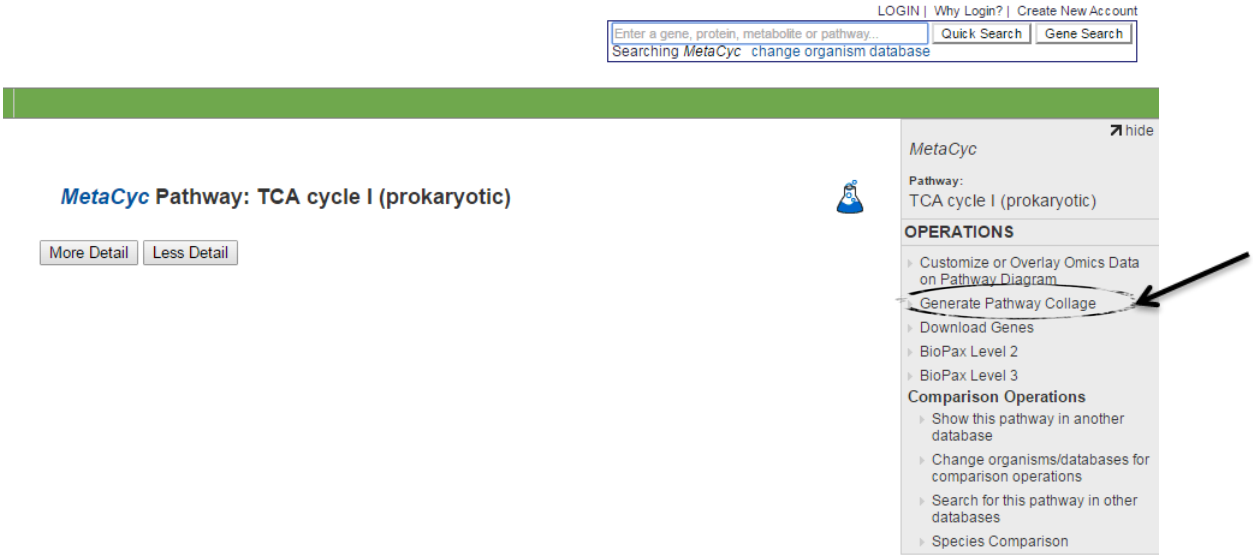
12.2. Na barra de pesquisa inserir: *TCA cycle* e clicar em ***Quick search*** (pesquisa rápida)

12.3. Selecionar: *TCA cycle I (prokaryotic)* no menu ***Pathways*** (Vias metabólicas)

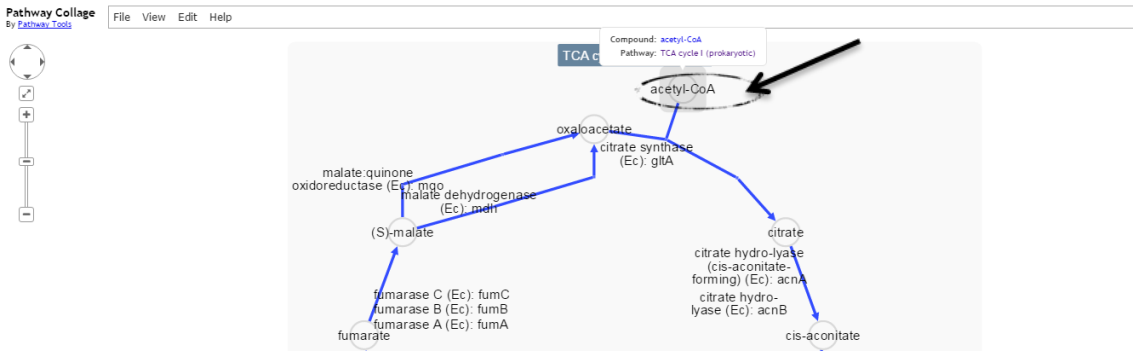




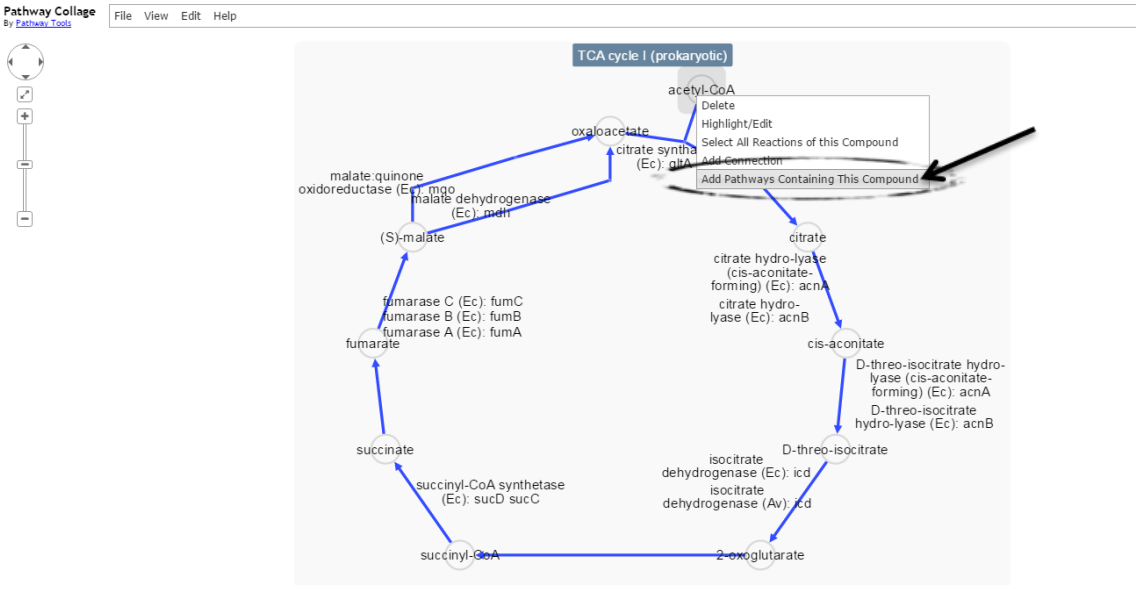
12.4. Após aceder à página com a descrição da reação, selecionar nas operações do lado direito do menu: ***Generate Pathway Collage***



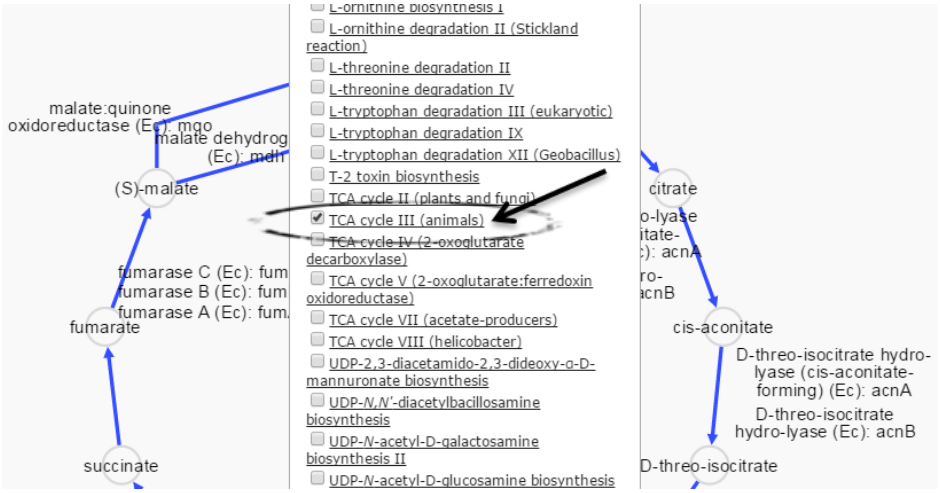
12.5. Ao clicar no nome do composto é possível obter informação sobre a via metabólica em que o composto intervém.



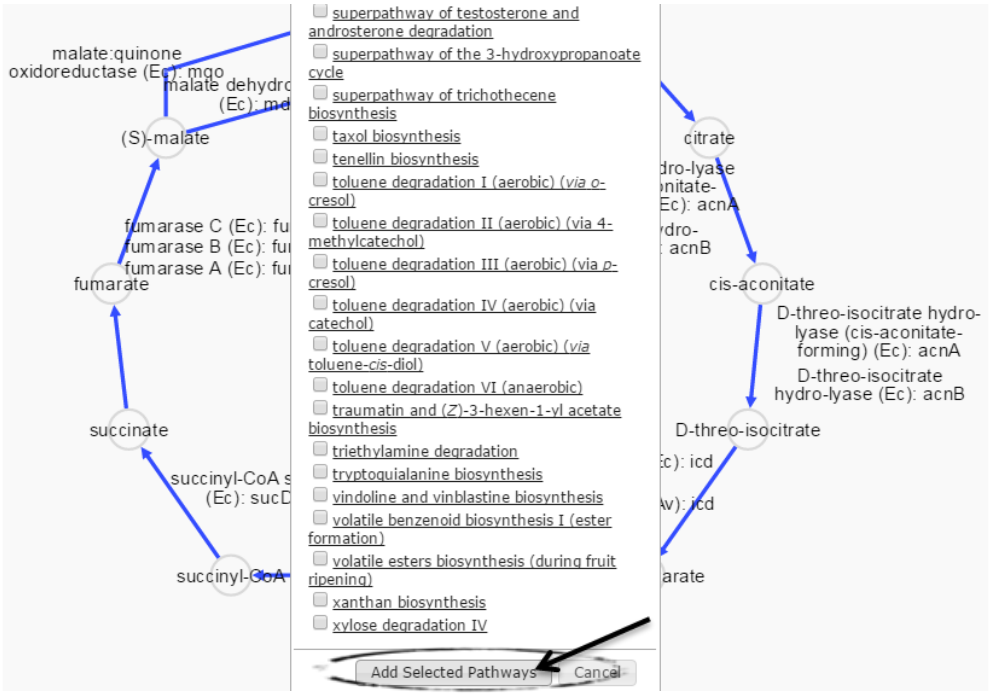
12.6. Clicar com o botão direito do rato em *acetyl-CoA* e escolher: "***Add pathways containing this compound***" (Adicionar vias metabólicas contendo este composto)



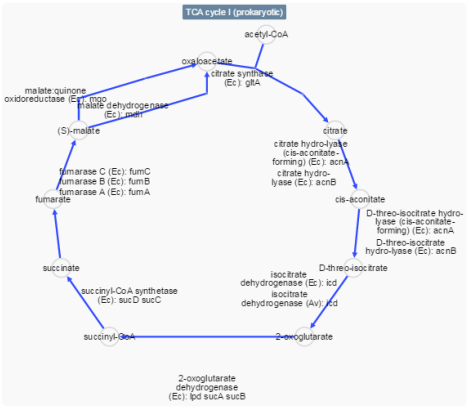
12.7. Selecionar ***TCA cycle III (Animals)***

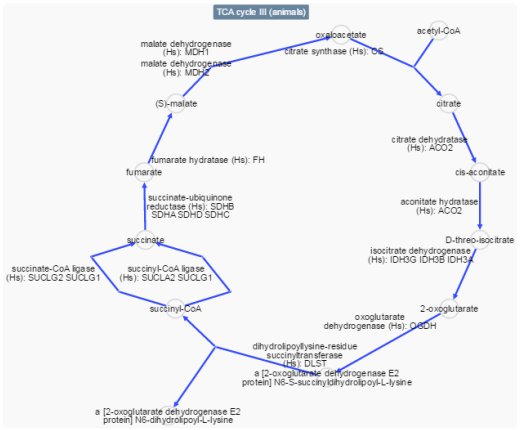


12.8. Escolher ***Add selected pathways*** (Adicionar vias metabólicas selecionadas)

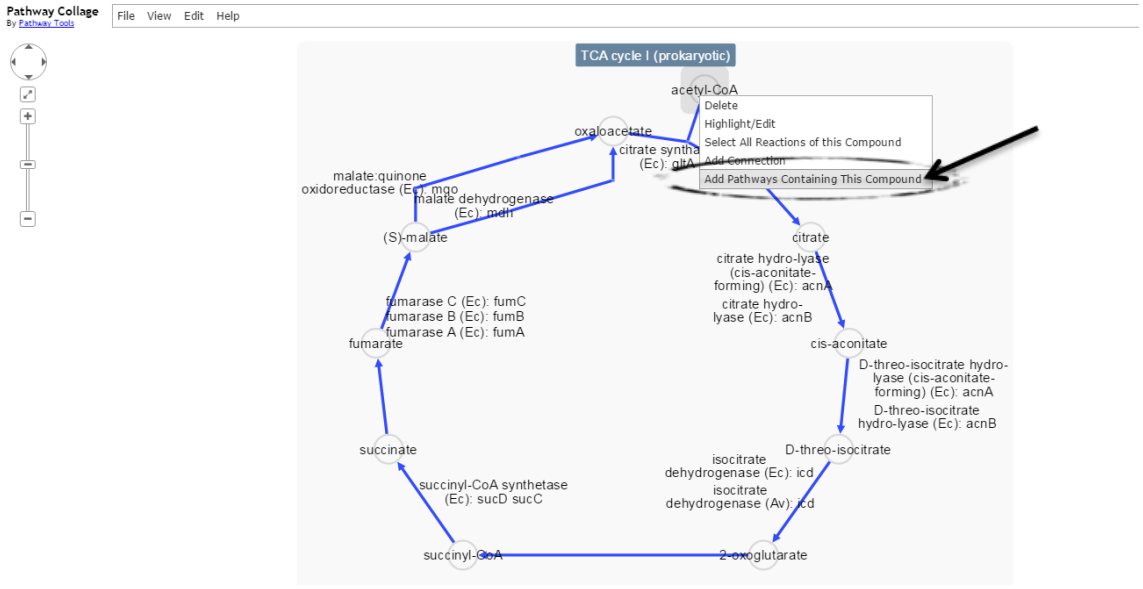


12.9. Comparar as semelhanças e as diferenças entre as moléculas intervenientes nos dois ciclos selecionados.

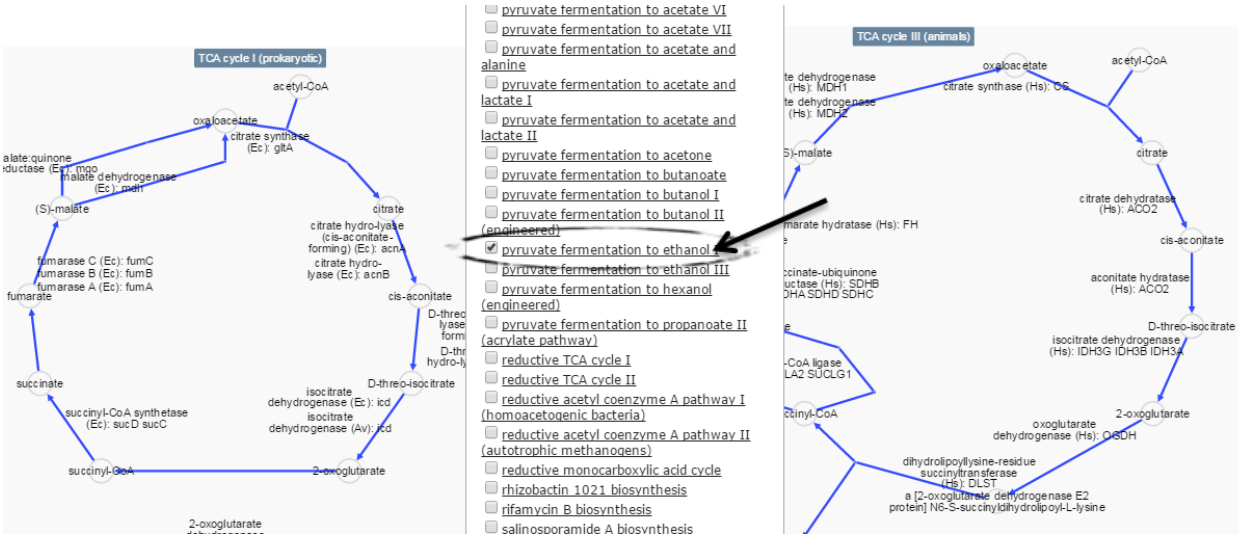




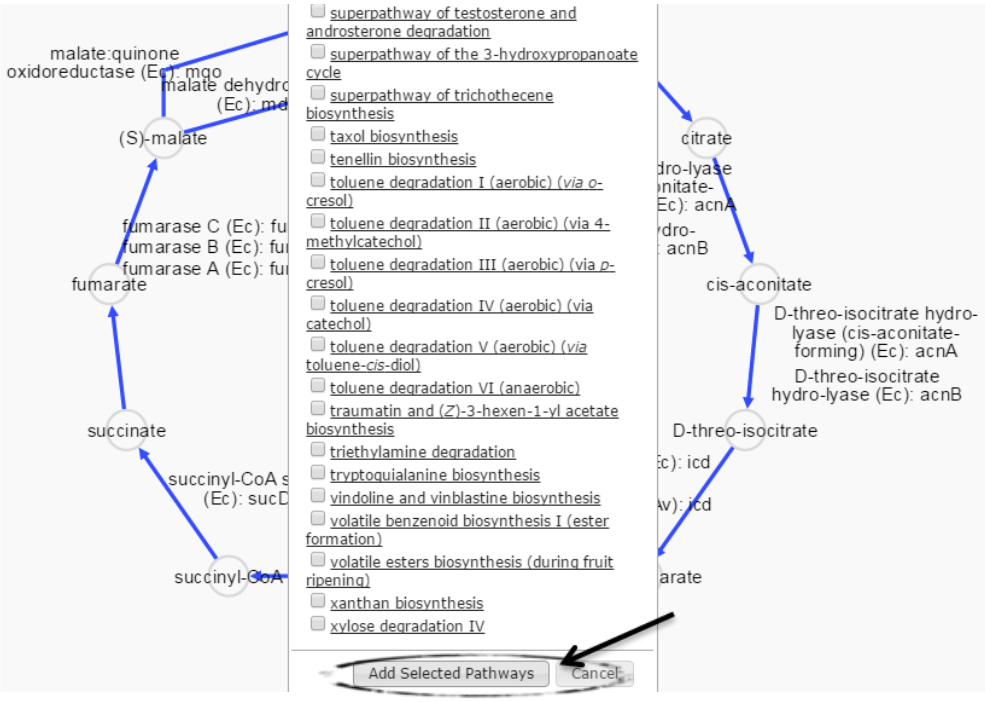
12.10. Selecionar o composto acetyl-CoA em um dos ciclos, clicar com o botão direto do rato e escolher: ***Add pathways containing this compound*** (Adicionar vias metabólicas contendo este composto)



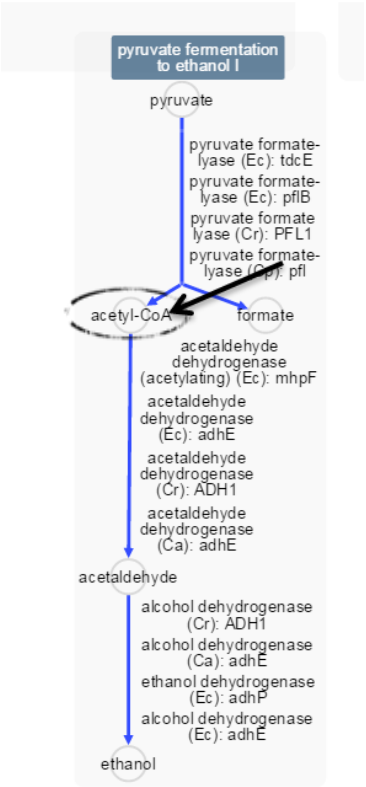
12.11. Selecionar: ***Pyruvate Fermentation to Ethanol I*** (Fermentação do Piruvato em Etanol I)



12.12. Escolher: ***Add selected pathways*** (Adicionar vias metabólicas selecionadas)

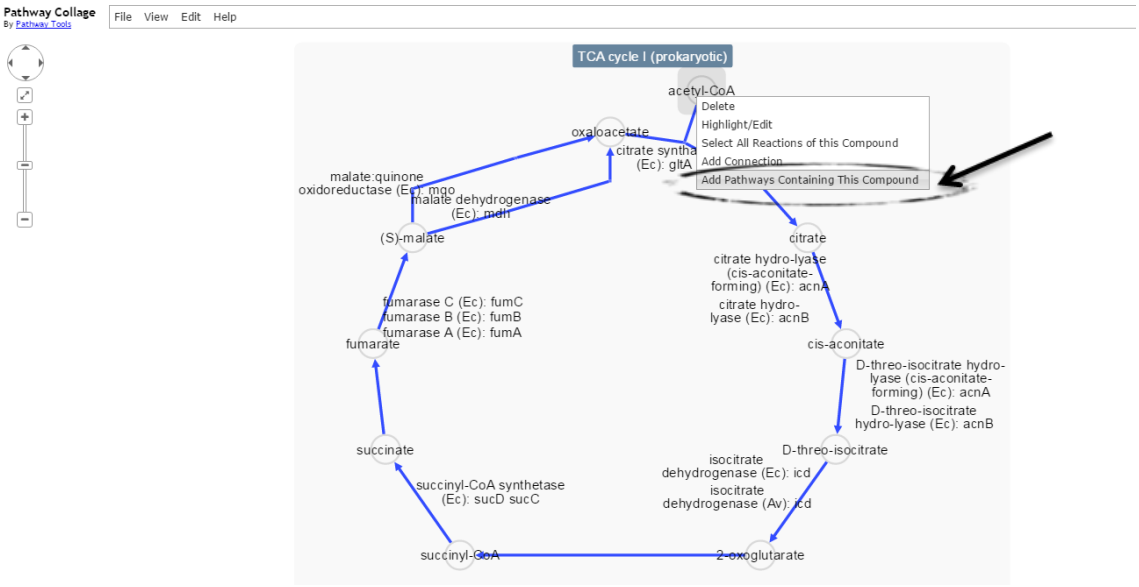


12.13. Localizar o composto acetyl-CoA na via metabólica adicionada anteriormente.

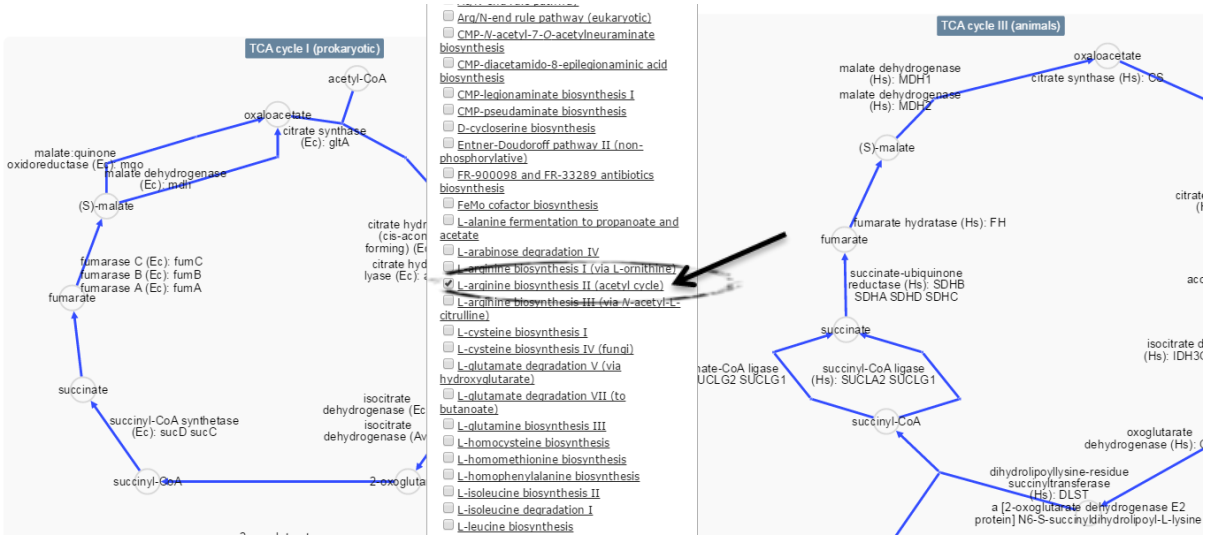


12.14. Concluir acerca da intervenção de um mesmo composto em diferentes vias metabólicas (exemplo: respiração aeróbia e fermentação).

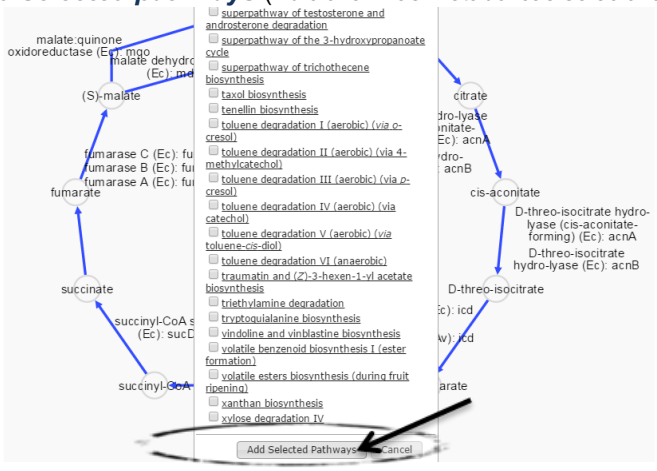
12.15. Selecionar o composto acetyl-CoA em um dos ciclos e escolher: ***Add pathways containing this compound*** (Adicionar vias metabólicas contendo este composto).



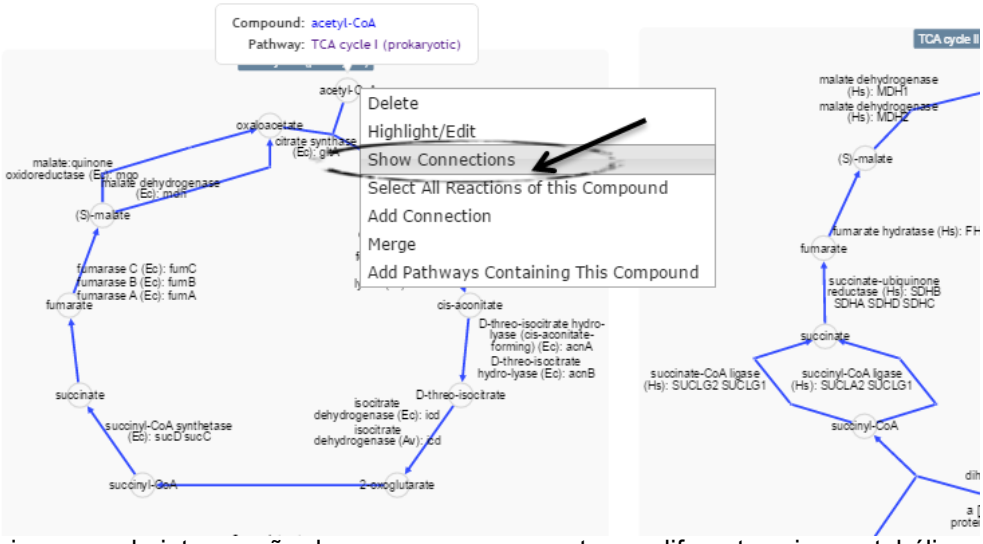
12.16. Selecionar: ***L-arginine biosynthesis II (acetyl cycle)***



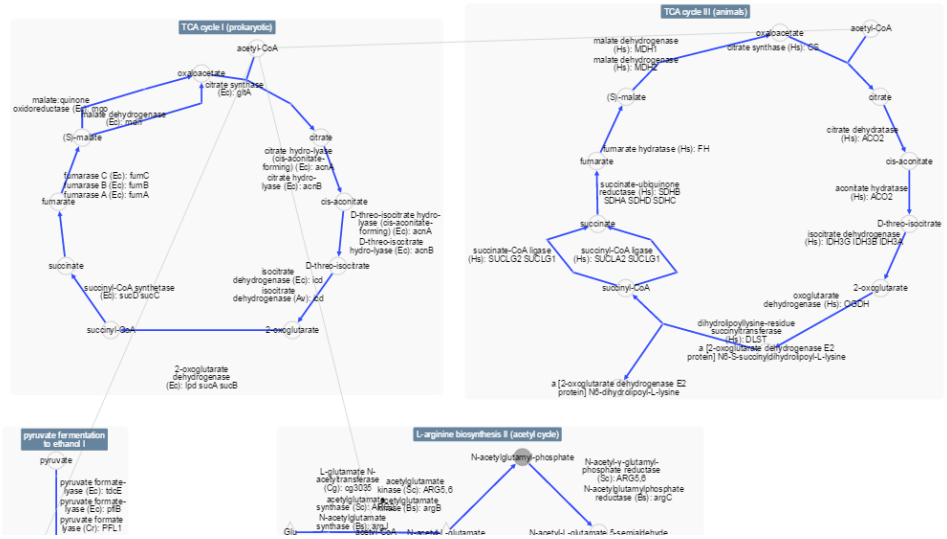
12.17. Selecionar: ***Add selected pathways*** (Adicionar vias metabólicas selecionadas)



12.18. Selecionar o composto acetyl-CoA e com o botão direito do rato e escolher: ***Add omitted side nodes for this compound*** e de seguida: ***Show all connections***.



12.19. Concluir acerca da intervenção de um mesmo composto em diferentes vias metabólicas (respiração aeróbia, fermentação e biossíntese de aminoácidos).



Explorando Vias Metabólicas nos Diferentes Domínios da Vida (Parte II)

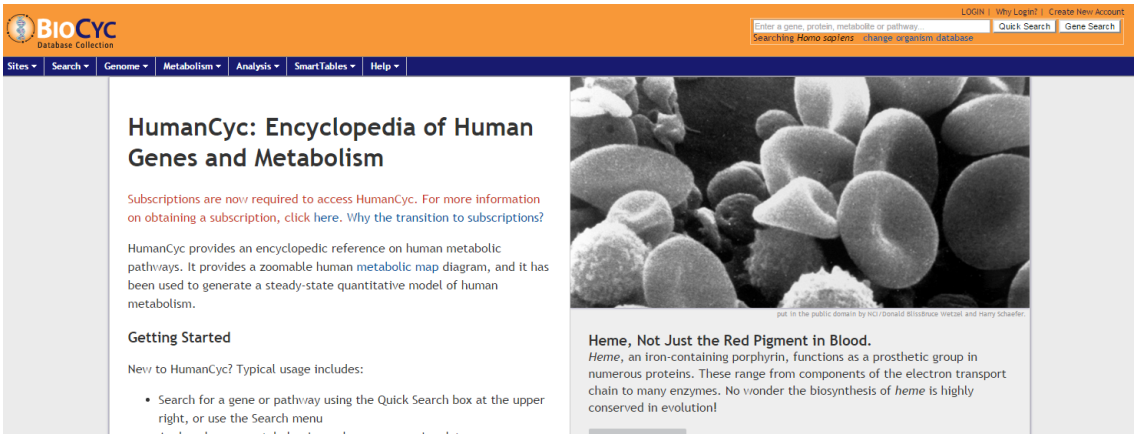
A plataforma bioinformática HumanCyc -*Encyclopedia of Human Genes and Metabolism* (http://humancyc.org/) permitirá estudar uma biomolécula presente no metabolismo humano – colesterol -, tornando possível analisar de forma detalhada a via metabólica que lhe dá origem, reconhecendo ainda outros compostos da mesma e que desempenham um papel crucial no organismo (exemplo: na produção de vitamina D). Este conhecimento, para além do seu valor científico, ao estabelecer a relação entre a síntese de diferentes biomoléculas, poderá ser um mobilizador da adoção de comportamentos mais saudáveis.

Enquadramento curricular:

Este exercício tem um enquadramento curricular vasto, a saber: 9ºano: Saúde individual e comunitária (2;2.5); 10º ano: Módulo Inicial: Diversidade na biosfera – 2. A célula; 2.2. Constituintes básicos; 12ºano: Módulo 2 – Controlo de doenças e biotecnologia: 1.2. Desequilíbrios e doenças.

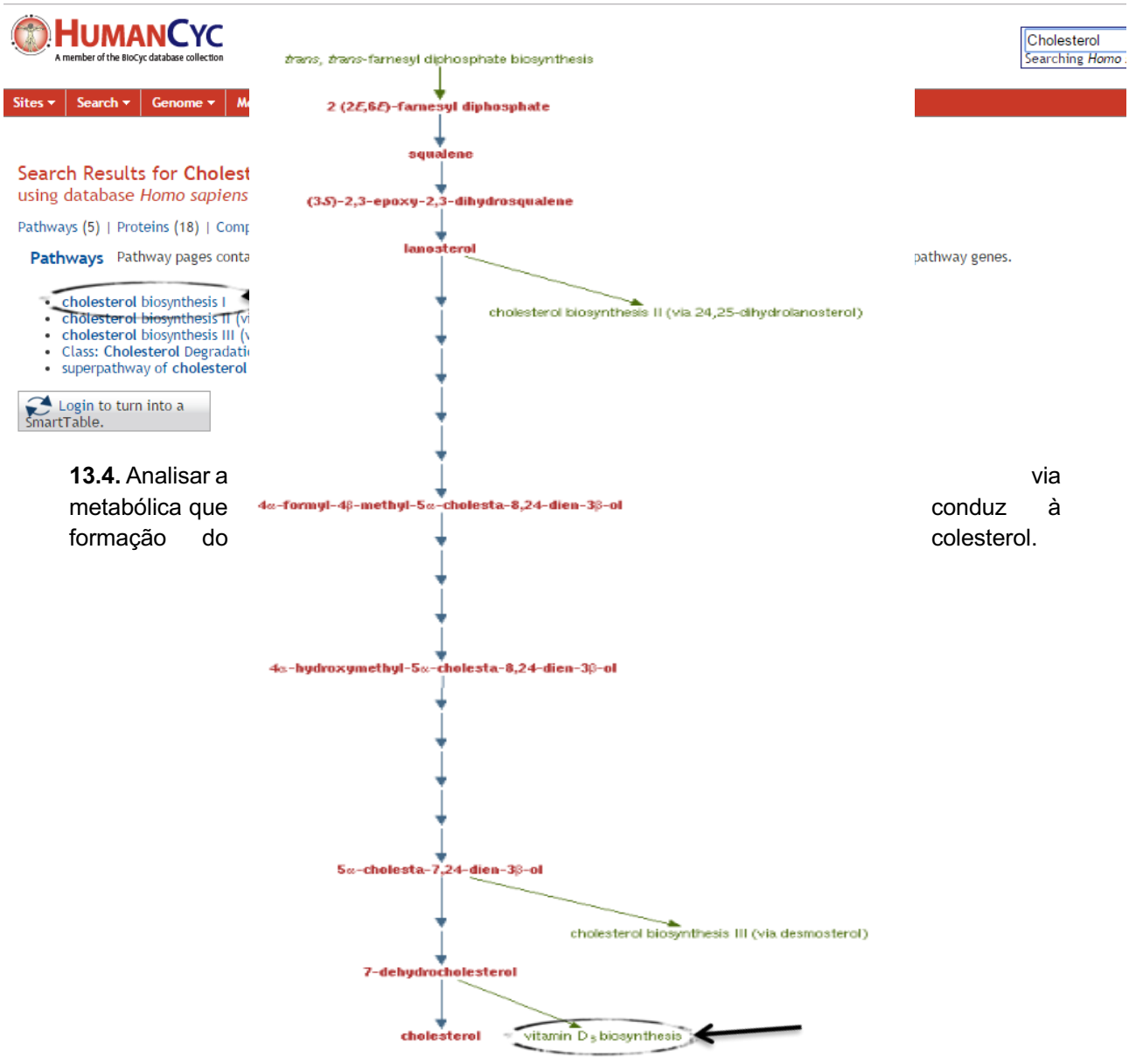
Exercício 13: *HumanCyc -Encyclopedia of Human Genes and Metabolism*

13.1. Aceder ao site: http://humancyc.org/



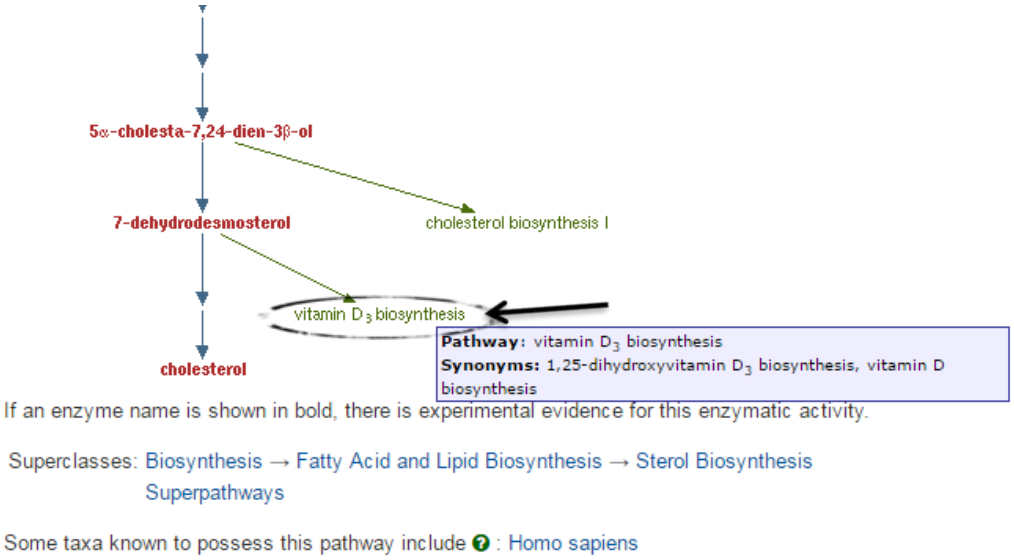
13.2. Na barra de pesquisa inserir o termo: *Cholesterol* e clicar em ***Quick search*** (Pesquisa Rápida)

13.3. Escolher a opção: ***Cholesterol biosynthesis I*** (Biossíntese de colesterol I)

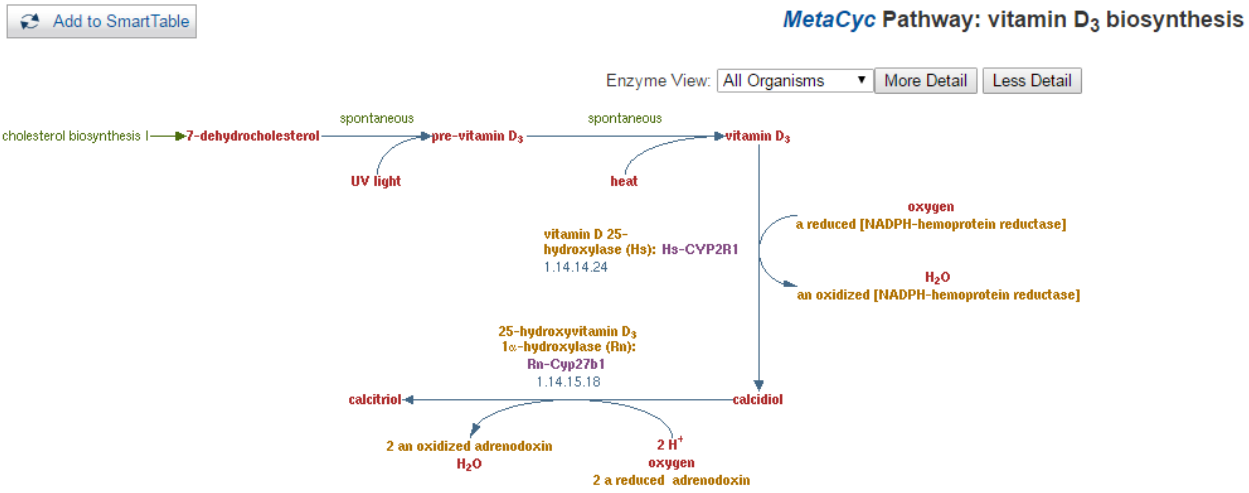


13.4. Analisar a via metabólica que conduz à formação do colesterol.

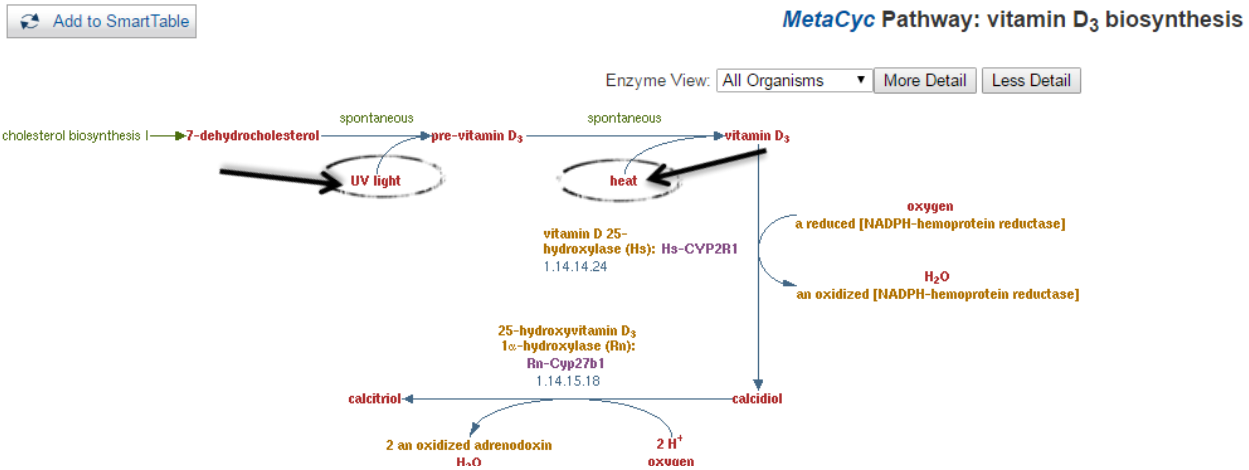
13.5. Clicar em ***Vitamin D3 biosynthesis*** *(Biossíntese de Vitamina D3)*



13.6. Concluir acerca da possibilidade de formação de outras biomoléculas através da mesma via metabólica (*Vitamin D3 biosynthesis*).



13.7. Identificar os fatores que interferem na via metabólica de produção desta vitamina D3.



13.8. Para pensar ...

A deficiência de vitamina D é um problema de saúde mundial. A principal fonte de vitamina D para a maioria dos seres humanos é a exposição solar. Vários fatores influenciam a produção cutânea de vitamina D como o uso de protetor solar, a pigmentação da pele, a hora do dia, a estação do ano, a latitude, e envelhecimento. A falta de exposição solar e a consequente deficiência de vitamina D têm sido associados a muitas doenças crónicas graves, incluindo doenças auto-imunes, doenças cardiovasculares e cancro. (Vitamin D and Sunlight: Strategies for Cancer Prevention and Other Health Benefits – 2008).

? De que forma é que a produção de vitamina D pode ser afetada em função das condições intervenientes na sua formação?

Exemplo: estações do ano mais frias - menor produção (o calor afeta esta reação); Menos insolação, menos produção (a radiação ultravioleta afeta esta reação).

? De que forma é que as populações de zonas geográficas com temperaturas médias anuais mais baixas e com baixa taxa de insolação podem manter níveis saudáveis de vitamina D?

A bioinformática ao serviço da população: exemplos práticos

(Parte I)

Utilizando a aplicação de análise evolutiva MEGA (*Molecular Evolutionary Genertics Analysis*) (http://www.megasoftware.net/), será feito um estudo sobre a evolução da resistência aos antibióticos (tetraciclina) em bactérias, que permita compreender melhor a emergência recente de estirpes de bactérias multirresistentes. O programa MEGA permite comparar diferentes organismos, analisando o fenómeno de evolução da resistência a antibióticos.

Enquadramento curricular:

Esta atividade enquadra-se nas metas curriculares propostas para o 9ºano de escolaridade no tema Saúde Individual e Comunitária: Resistência a antibióticos (1;1.6.). Serão ainda explorados os eventos de transferência vertical e lateral de genes.

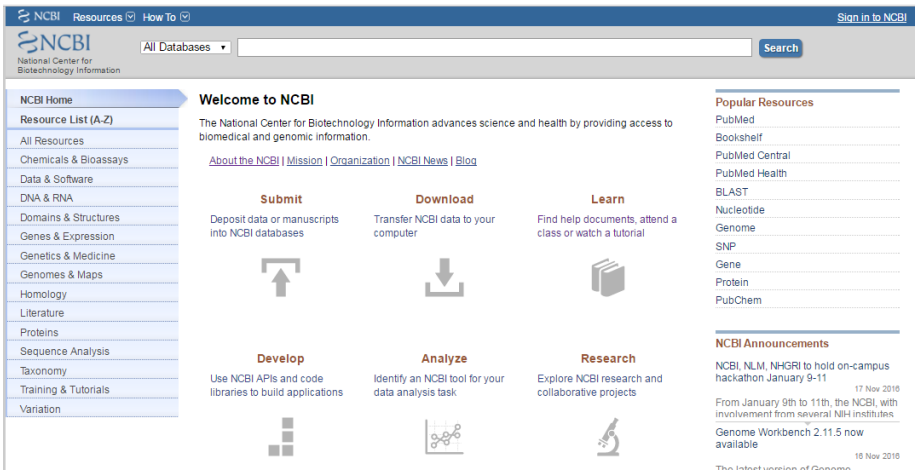
Exercício 14: *Evolução e resistência a antibióticos*

A seguinte proteína é responsável pela resistência ao antibiótico tetraciclina no organismo *Ureaplasma urealyticum*:

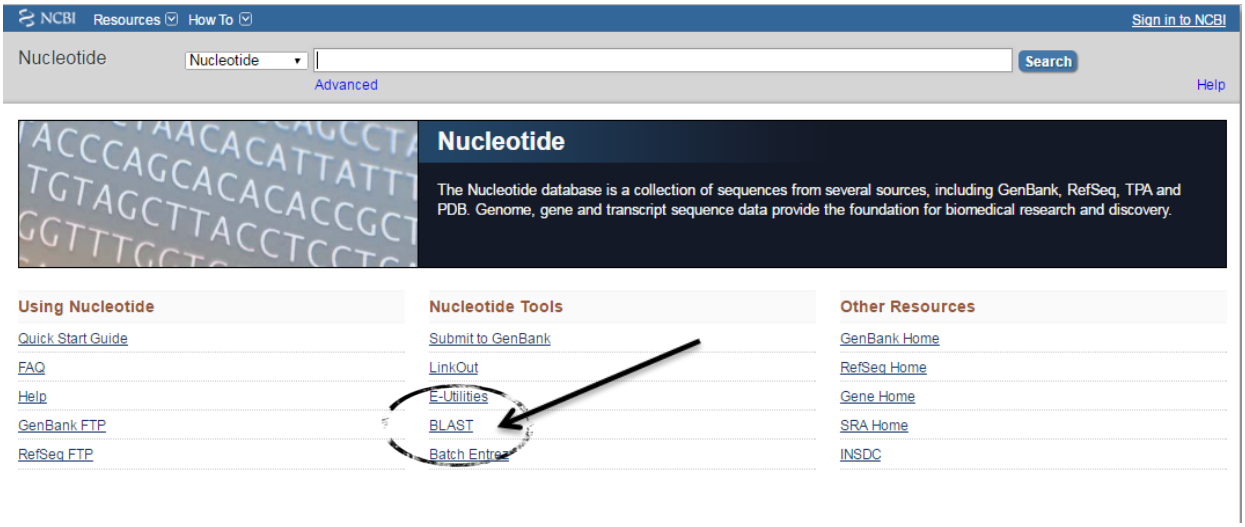
>gi|475984|gb|AAA73978.1| tetracycline resistance determinant [Ureaplasma urealyticum]

MKIINIGVLAHVDAGKTTLTESLLYNSGAITELGSVDKGTTRTDNTLLERQRGITIQTGITSFQWENTKVNIIDT
PGHMDFLAEVYRSLSVLDGAILLISAKDGVQAQTRILFHALRKMGIPTIFFINKIDQNGIDLSTVYQDIKEKLSA
EIVIKQKVELYPNMCVTNFTESEQWDTVIEGNDDLLEKYMSGKSLEALELEQEESIRFHNCSLFPVYHGSA
KNNIGIDNLIEVITNKFYSSTHRGPSELCGNVFKIEYTKKRQRLAYIRLYSGVLHLRDSVRVSEKEKIKVTEMY
TSINGELCKIDRAYSGEIVILQNEFLKLNSVLGDTKLLPQRKRIENPHPLLQITVEPSKPEQREMLLDALLEIS
DSDPLLRYYVDSTTHEIILSFLGKVQMEVISALLQEKYHVEIELKEPTVIYMERPLKNAEYTIHIEVPPNPFWA
SIGLSVSPLPLGSGMQYESSVSLGYLNQSFQNAVMEGIRYGCEQGLYGWNVTECKICFKYGLYYSPVSTP
ADFRMLAPIVLEQVLKKAGTELLEPYLSFKIYAPQEYLSRAYNDAPKYCANIVDTQLKNNEVILSGEIPARCI
QEYRSDLTFFTNGRSVCLTELKGYHVTTGEPVCQPRRPNSRIDKVRYMFNKIT

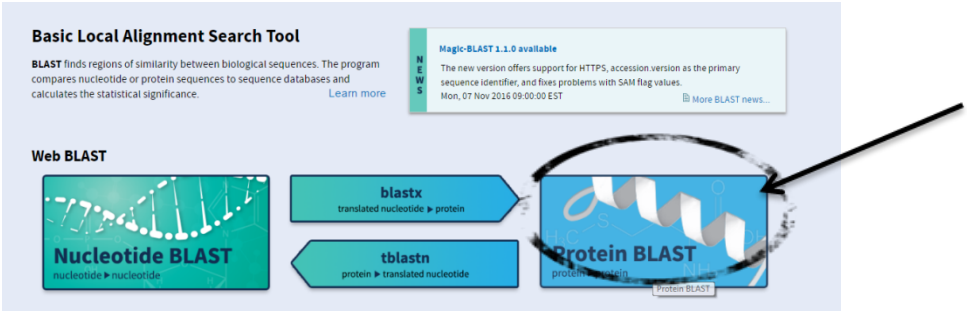
14.1. Aceder à página do National Center for Biotechnology Information (NCBI): https://www.ncbi.nlm.nih.gov/

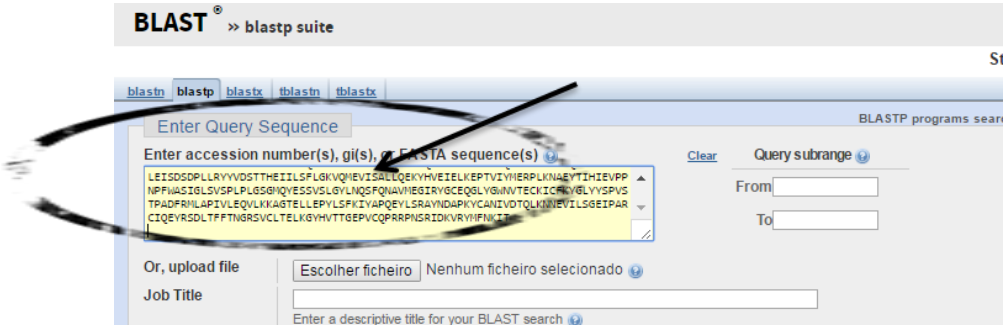


14.2. Aceder à ferramenta ***BLAST.***



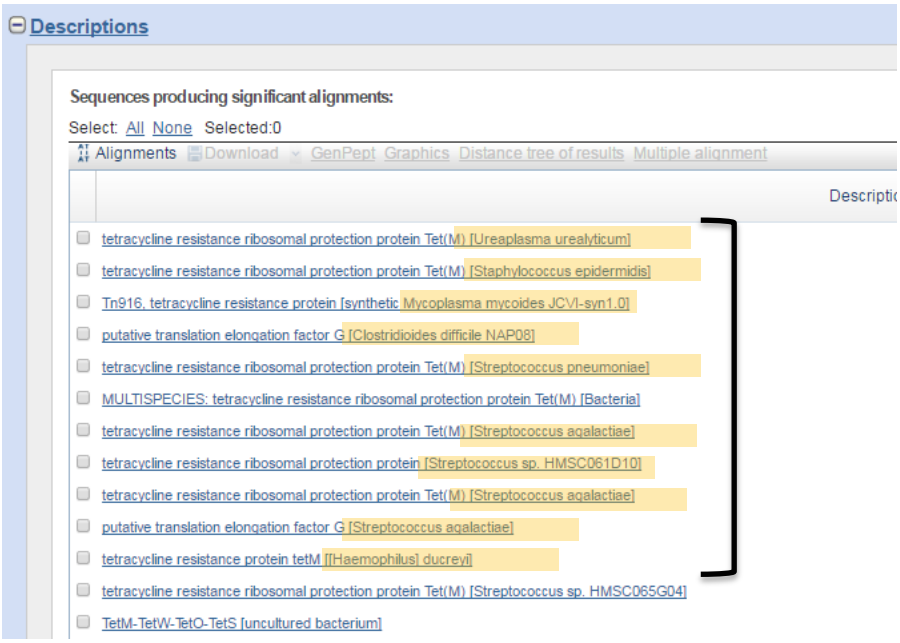
14.3. Selecionar *"**protein blast**"* (BLAST de proteínas), e fazer *copy-paste* da sequência em estudo para o campo *"**query sequence**"*.



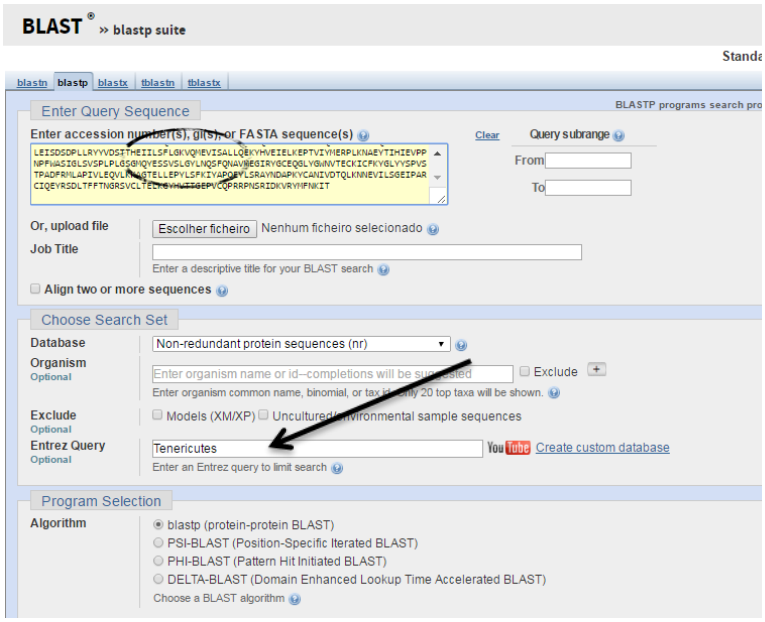


14.4. Iniciar a busca clicando em "***BLAST***".

14.5. Verificar a que organismos pertencem as 10 proteínas mais semelhantes à proteína em estudo.

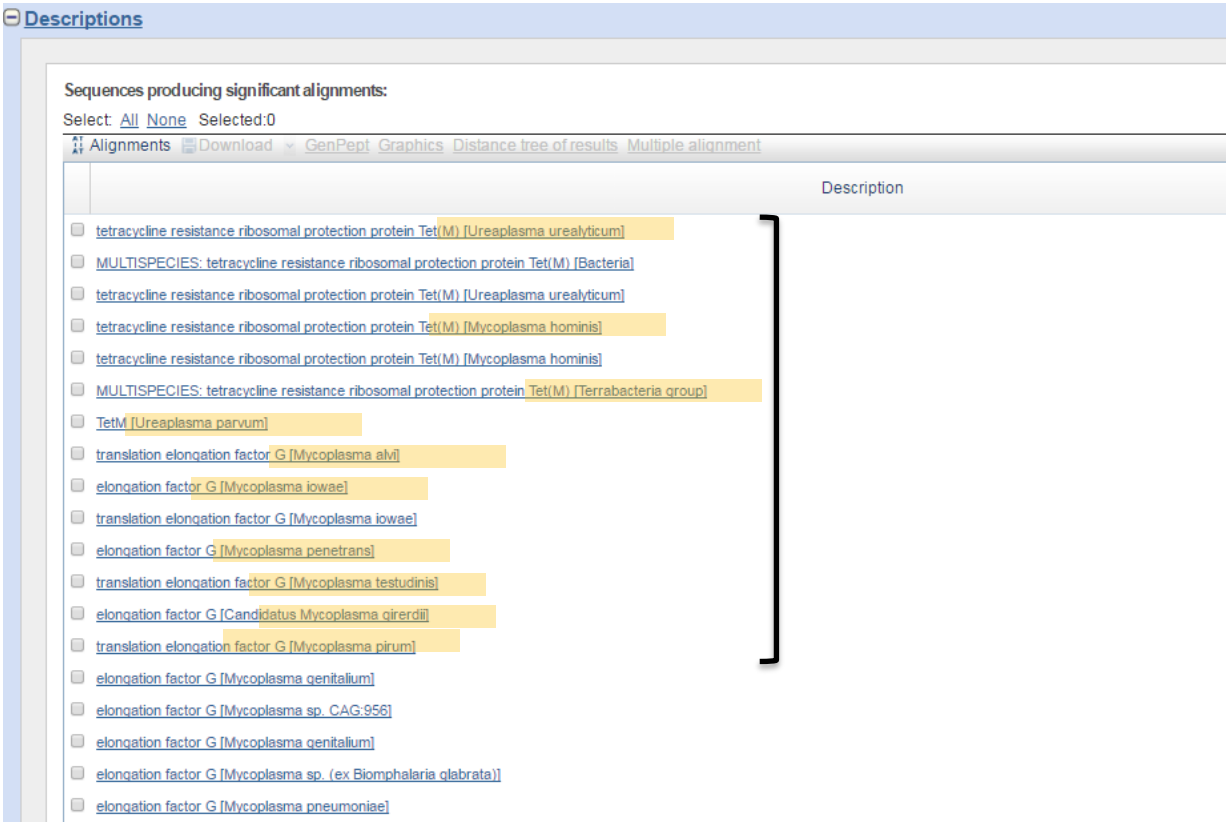


14.6. Voltar ao formulário do Blast e escrever "*Tenericutes*" no campo "***Entrez query***" (dessa forma, a pesquisa ficará limitadas ao filo *Tenericutes*, ao qual pertence *Ureaplasma urealyticum*).



14.7. Fazer nova busca clicando em "***BLAST***".

14.8. Verificar quais os organismos, dentro dos Tenericutes, a pertencem as 10 proteínas mais semelhantes à proteína em estudo.



14.9. Comparar o primeiro resultado com o segundo. O que se pode concluir?

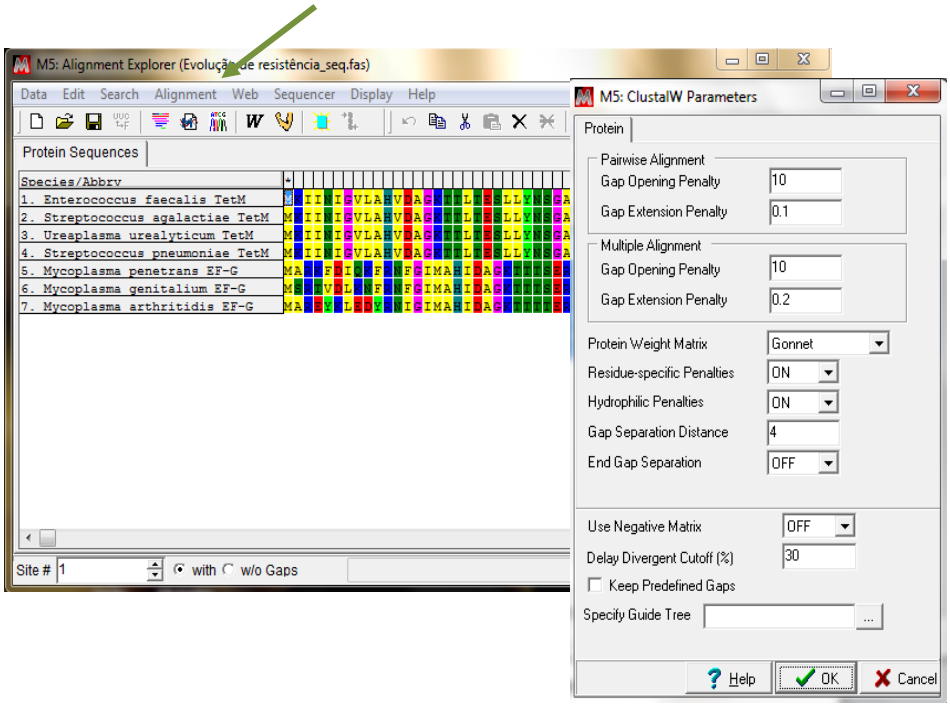
As proteínas mais próximas da proteína considerada inicialmente não se encontram entre os organismos aparentados de *U. urealyticum* – a identidade e as relações de similaridade de cada gene em particular não seguem obrigatoriamente aquelas do organismo a que pertence.

14.10. No computador, abrir o ficheiro seq.fas utilizando o programa Mega (fazer duplo click sobre o ficheiro). P*ara mais informações sobre este programa, carregue aqui.*

Este ficheiro contém a proteína original, TetM de *U. urealitycum*, que faz parte do mecanismo de resistência ao antibiótico tetraciclina, bem como as proteínas homólogas (isto é, com a mesma origem) elongation-factor G. Estas proteínas EF-G são responsáveis pela translocação do mRNA e dos tRNAs ao longo dosribossomas durante a síntese proteica.

A tetraciclina exerce a sua acção antibiótica impedindo a ligação entre os tRNAs e os ribossomas: dessa forma, a produção proteica cessa e o organismo acaba por morrer. As proteínas TetM funcionam como protectoras dos ribossomas, possibilitando a síntese proteica mesmo na presença de tetraciclina.

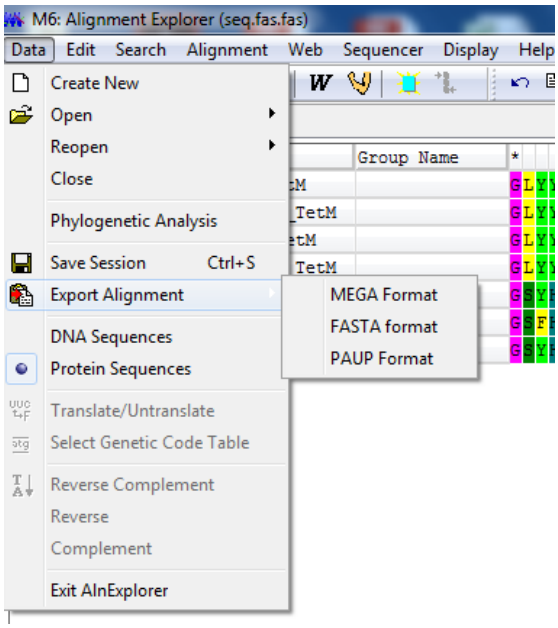
14.11. Alinhar as sequências (seleccionando todas as sequências no menu **Alignment** seleccionar a primeira opção *Align by ClustalW/* na janela de opções que aparece, manter as opções pré-definidas e clicar em **OK**).



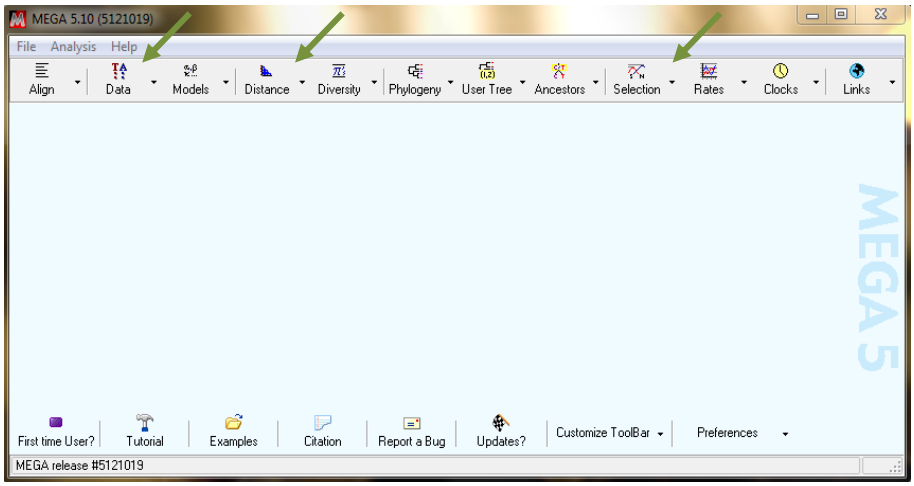
14.12. Observar o alinhamento. É possível verificar que há zonas completamente alinhadas (conservadas) entre todas as sequências, zonas conservadas entre algumas sequências, e zonas que são completamente diferentes em todas as sequências.

<u>O que é que a conservação de cada zona específica da proteína diz sobre a funcionalidade dessas regiões?</u>

14.13. Exportar o alinhamento para o formato .mega (seleccionando a opção *Export Alignment>MEGA format* no menu **Data**).

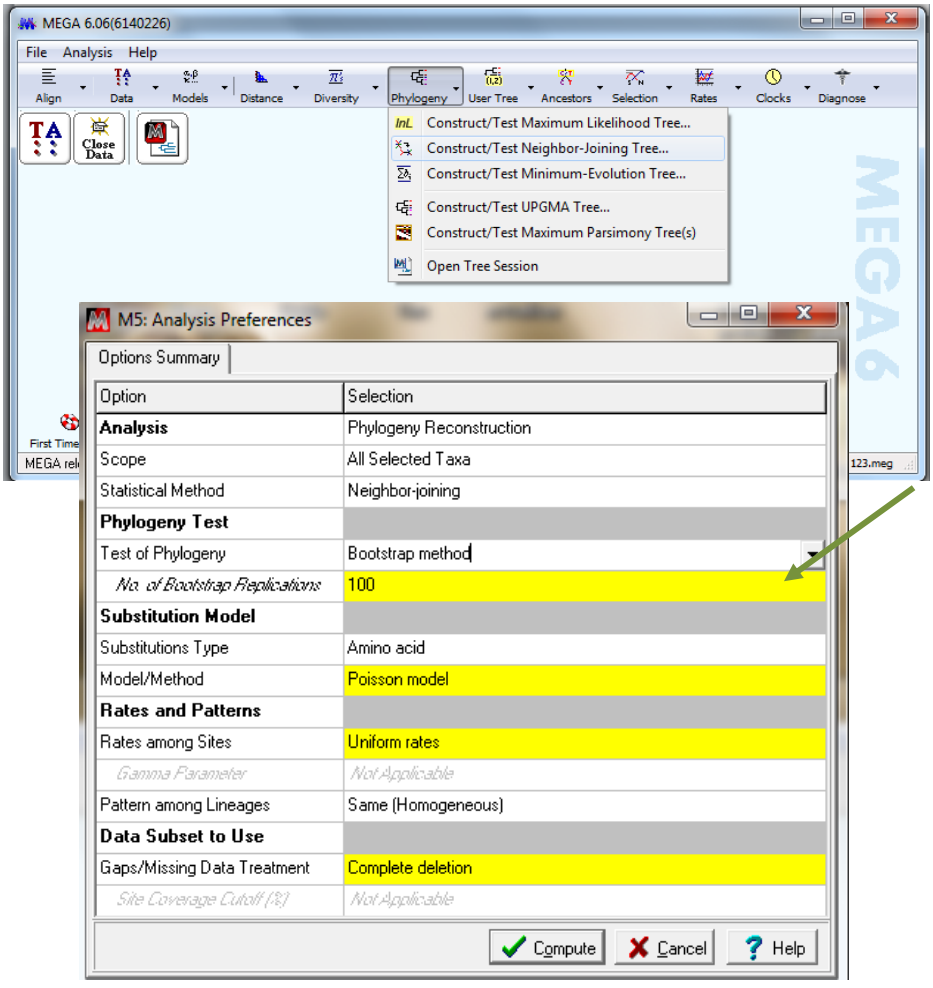


14.14. Fechar a janela do alinhamento e abrir o ficheiro exportado para .mega. Verificar o aparecimento de um conjunto de novas opções na janela inicial do Mega.



14.15. Construir uma árvore evolutiva a partir do alinhamento das sequências pelo método Neighbor-Joining (no menu **Phylogeny**, seleccionar a opção *Construct/Test >Neighbor-Joining tree*).

Na janela de opções que surge, certificar-se que na opção *Phylogeny Test*, está definido *Bootstrap* com 100 réplicas.



14.16. Observar a árvore construída. Com base nos resultados obtidos e sabendo que:

1) *Enterococcus faecalis*, *Streptococcus agalactiae* e *Streptococcus pneumoniae* pertencem ao filo **Firmicutes**;

2) *Mycoplasma penetrans*, *Mycoplasma genitalium*, *Mycoplasma arthritidis* e *Ureaplasma urealyticum* pertencem ao filo **Tenericutes**;

Que conclusões se podem retirar sobre a evolução da proteína TetM na espécie *Ureaplasma urealyticum*?

14.17. Embora *Ureaplasma urealyticum* faça parte da flora normal do tracto urogenital dos humanos, pode, sob determinadas condições (por exemplo, imunossupressão medicamentosa ou outra) causar infeções urinárias.

Quais são as consequências da aquisição por parte destes microrganismos do gene *TetM*, nomeadamente no que diz respeito ao tratamento destas infeções?

Para pensar…

- Comentar a seguinte frase:

Os mecanismos de resistência aos antibióticos são mais variados e estão mais disseminados do que a produção de antibióticos em si, que é restrita a um pequeno grupo de microrganismos.

Tendo em consideração que a produção de antibióticos para os microrganismos produtores, representa:

1) uma vantagem evolutiva (competição com outras populações bacterianas pelo mesmo nicho ecológico);

2) um custo metabólico elevado (em termos de matéria prima e energia).

A bioinformática ao serviço da população: exemplos práticos (Parte II)

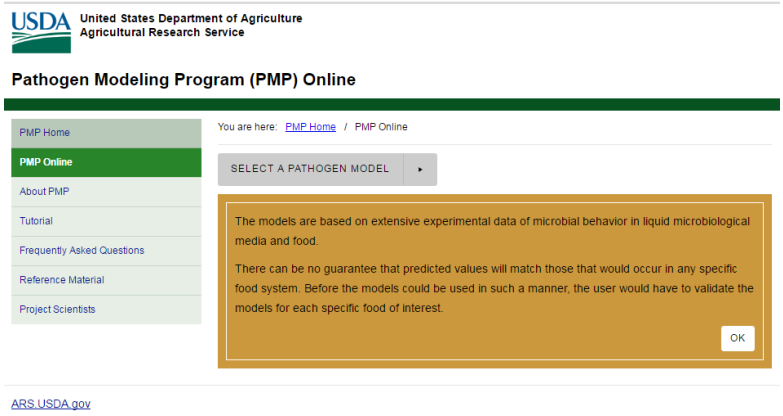
Serão exploradas as ferramentas disponibilizadas no *Pathogen Modeling Program (PMP) Online* (http://pmp.errc.ars.usda.gov/PMPOnline.aspx) e no *ComBase* (http://www.combase.cc/index.php/en/) tendo como principais objetivos: identificar as valias de diferentes métodos de preservação de alimentos, e compreender o efeito de fatores extrínsecos (como a temperatura e concentração de oxigénio); de fatores intrínsecos (como o pH e osmolaridade); e de alguns aditivos alimentares. Esta atividade é ainda um ótimo paradigma para se compreender a importância de modelos preditivos para a indústria alimentar, sistematizados através de aplicações computacionais.

Enquadramento curricular:

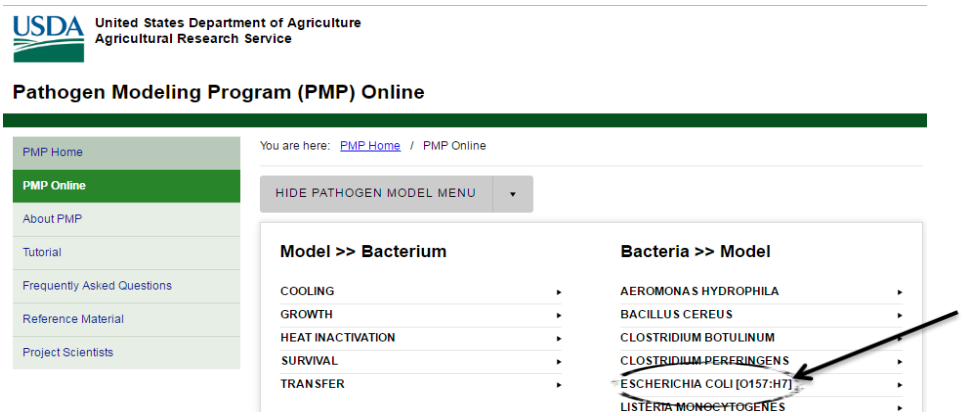
Esta atividade tem a particularidade de se enquadrar em dois ciclos de ensino diferentes: 8ºano de escolaridade: Sustentabilidade na Terra – Gestão sustentável dos recursos – 18. Relacionar o desenvolvimento científico e tecnológico com a melhoria da qualidade de vida das populações humanas; e 12ºano: Módulo 2 – Controlo de Doenças e Biotecnologia: 3 – Microrganismos e Indústria Alimentar.

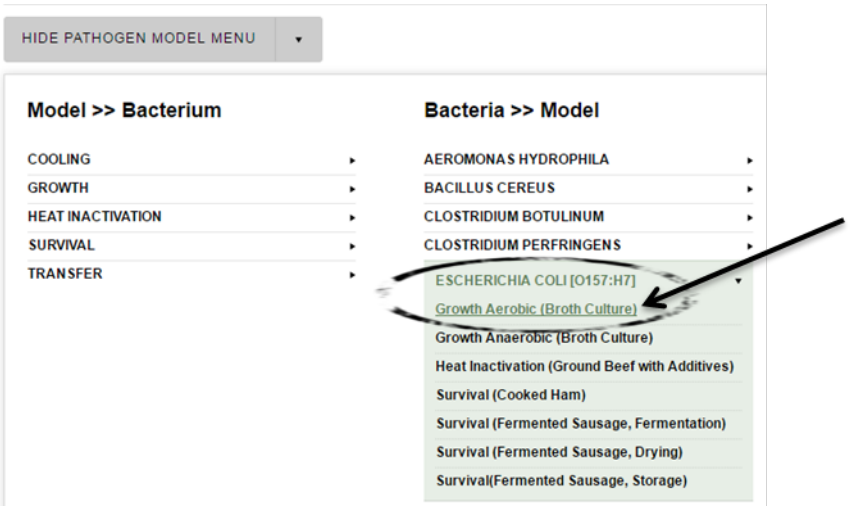
Exercício 15: *Pathogen Modeling Program (PMP) – modelos de crescimento*

15.1. Aceder à plataforma digital *Pathogen Modeling Program*: http://pmp.errc.ars.usda.gov/PMPOnline.aspx

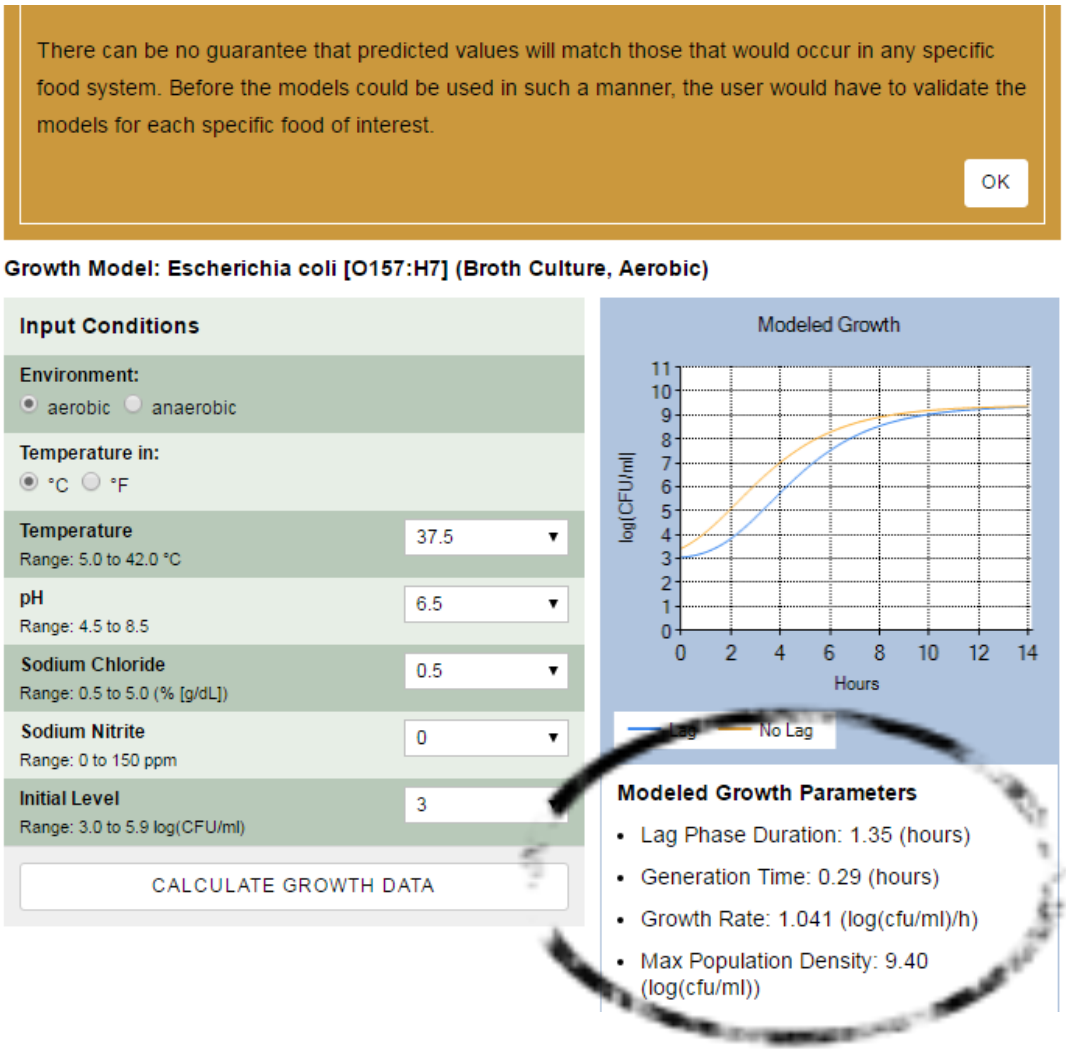


15.2. Abrir o menu *"Select a Pathoge Model"* e selecionar na secção *Bacteria >> Model* ***Escherichia coli [O157:H7]*** modelo ***Growth Aerobic Broth Culture.***

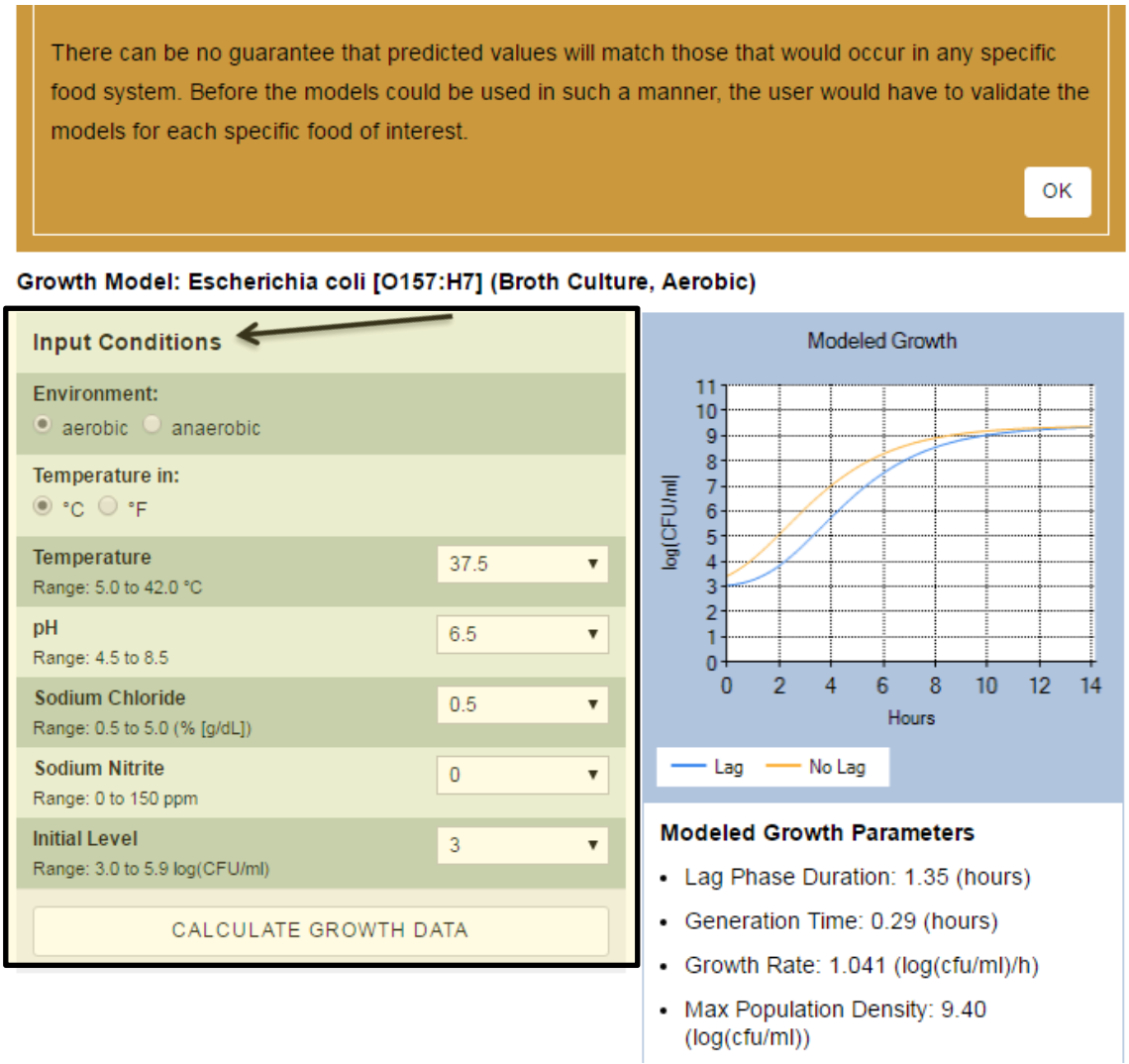




15.3. Identificar as diferentes fases de crescimento bacteriano e analisar os seus parâmetros, nomeadamente o tempo de duplicação e a taxa de crescimento.



15.4. Alterar as condições em que a *Escherichia coli* se encontra (à exceção do nitrito de sódio – *sodium nitrite*), e responder às questões a, b e c, justificando com dados recolhidos na plataforma, nomeadamente o tempo de duplicação (***Generation time***) e a taxa de crescimento (***Growth Model)***:



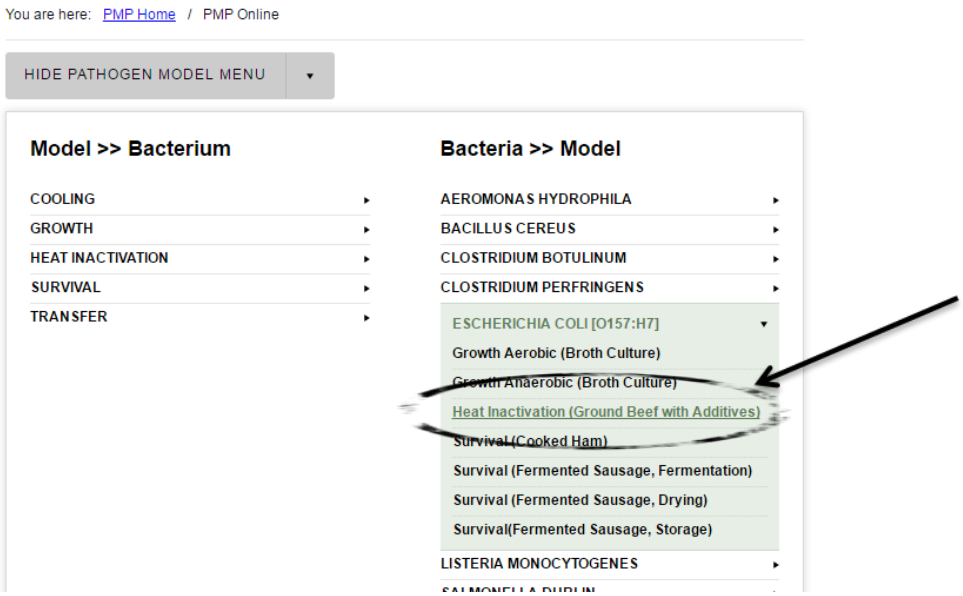
a) Comparar o crescimento destes microrganismos em condições de aerobiose e de anaerobiose. O que se pode concluir?

b) Verificar que para temperaturas inferiores a 10° C, *Escherichia coli* cresce lentamente (tem um longo tempo de duplicação).

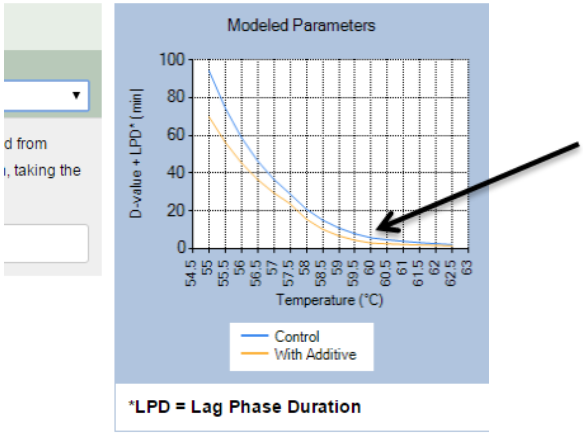
c) Sabendo que temperaturas de 10° C ou inferiores implicam um enorme consumo de energia, tornando o processo de preservação de alimentos muito dispendioso, explorar outros fatores que possam ser alterados (por exemplo, pH e cloreto de sódio) de modo a que, mantendo os alimentos à temperatura ambiente (20° C), as culturas de *Escherichia coli* mantenham um crescimento igual ou inferior ao obtido a 10° C.

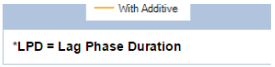
Exercício 16: *Pathogen Modeling Program (PMP) – inativação pelo calor*

16.1. Abrir o menu *"Select a Pathogen Model"* e selecionar na secção *Bacteria >> Model* ***Escherichia coli [O157:H7]*** modelo ***Heat Inactivation (Ground Beef with Additives).***



16.2. Analisar o tempo necessário para destruir estes microrganismos por inativação térmica, a 60 ºC, sem aditivo.





MODELED PARAMETERS

Temperature (°C)	Control D-Value + LPD (min)	Additive D-Value + LPD (min)
55.0	94.31	69.89
55.5	74.46	56.31
56.0	58.79	45.36
56.5	46.42	36.55
57.0	36.65	29.44
57.5	28.94	23.72
58.0	20.90	15.57
58.5	15.09	10.22
59.0	10.89	6.71
59.5	7.87	4.40
60.0	5.68	2.89
60.5	4.65	2.51
61.0	3.81	2.19

Para pensar …

i) Considerando que o grupo de empresas em questão, necessita de um processamento de inativação de *Escherichia coli* mais rápido do que o obtido, sugerir, entre as várias hipóteses apresentadas no quadro, a forma mais eficaz de atuação, sem recorrer ao uso de aditivos alimentares.

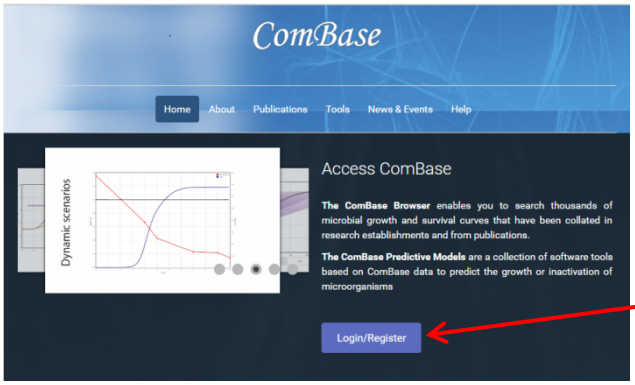
ii) Considerando a utilização de aditivos alimentares, refletir sobre o efeito dos aditivos alimentares (disponíveis na plataforma) na inativação térmica destas bactérias para valores de temperatura de 60 ºC (sem aditivos alimentares).

Exercício 17: *ComBase – modelos de crescimento*

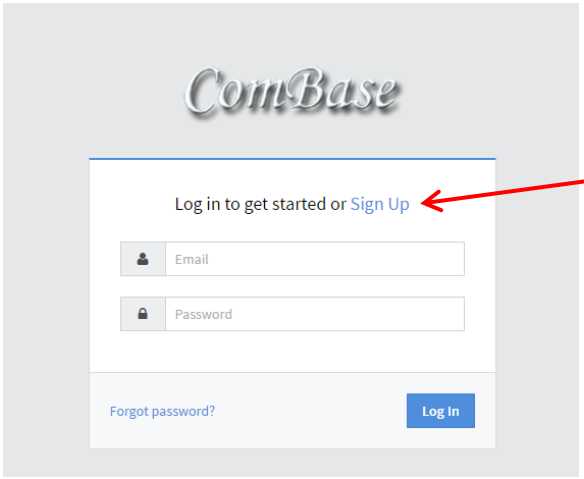
17.1. Aceder à plataforma digital *ComBase*: http://www.combase.cc/index.php/en/

Nota: Na primeira utilização da plataforma é necessário fazer o registo. Para isso:

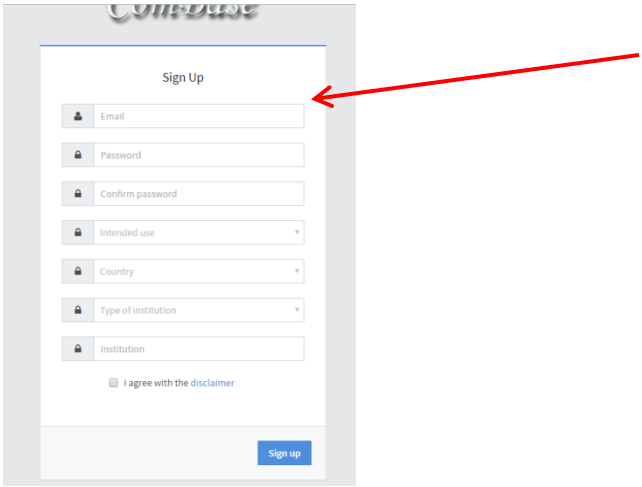
17.1. Fazer o *Login/Register*



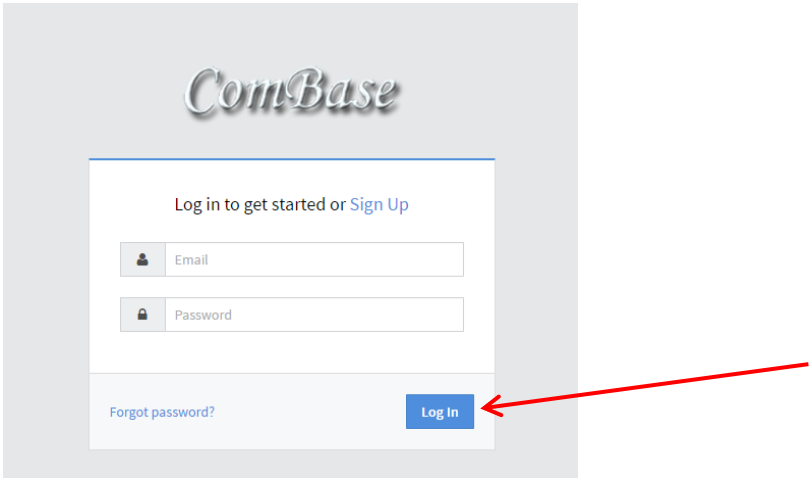
17.2. Clicar em Sign Up



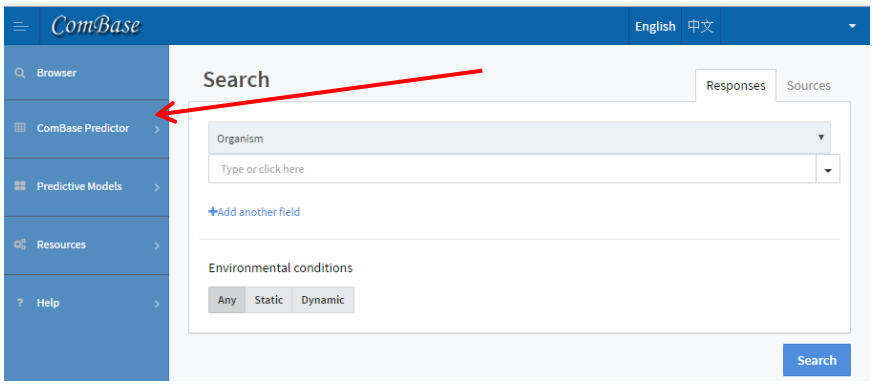
17.3. Efetuar o registo:



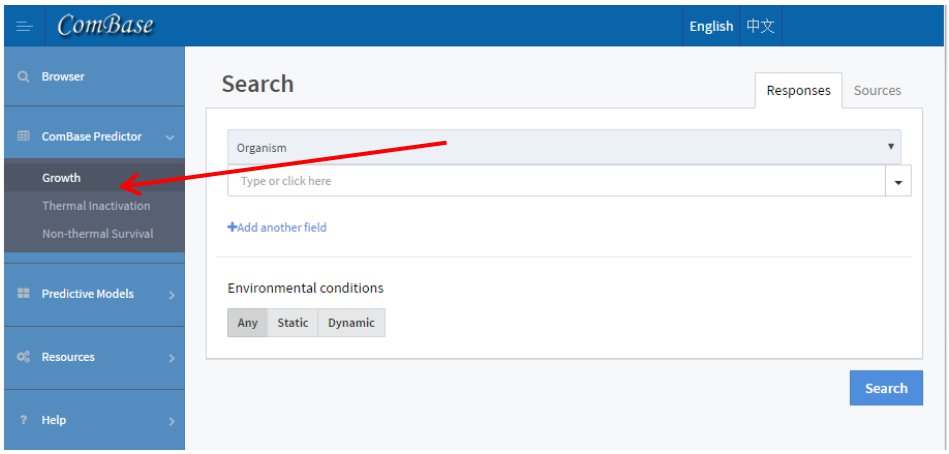
17.4. Fazer o *Log in* na plataforma com os dados de registo (recebidos no email):



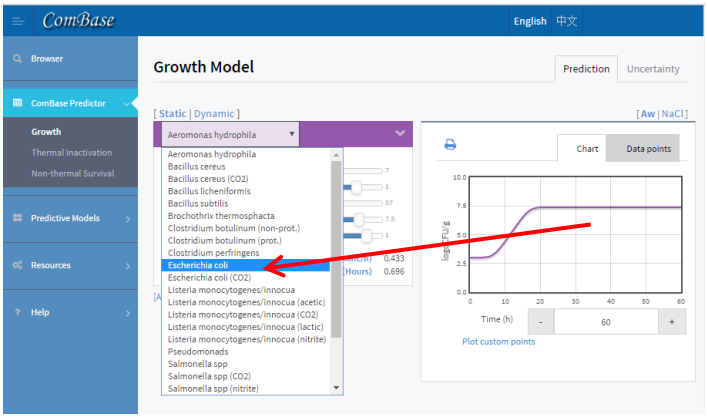
17.5. Aceder ao menu *ComBase Predictor*



17.6. Selecionar o modelo de crescimento (*Growth*)

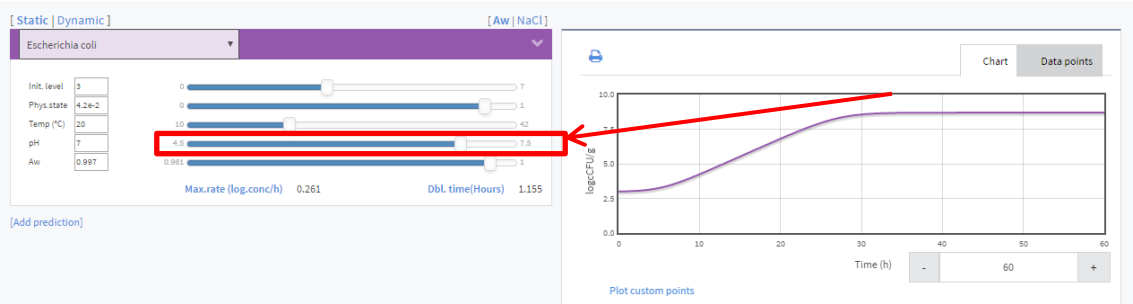


17.7. No organismo a estudar escolher *Escherichia coli*



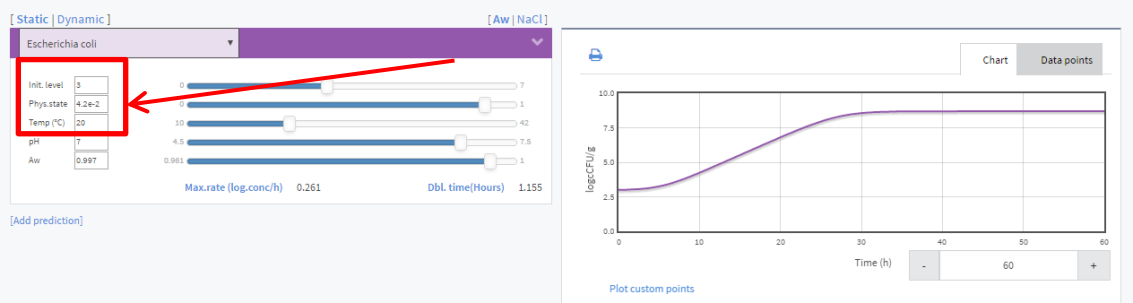
17.8. Identificar as diferentes fases de crescimento bacteriano e analisar os parâmetros: Max. *Rate* e *Dbl.time*

17.9. Alterar as condições em que a *Escherichia coli* se encontra e responder às questões a e



b, justificando com dados recolhidos na plataforma, nomeadamente o Max. Rate e o Dbl.time:

a) Para temperaturas inferiores a 10˚C, o crescimento da bactéria é lento ou rápido?

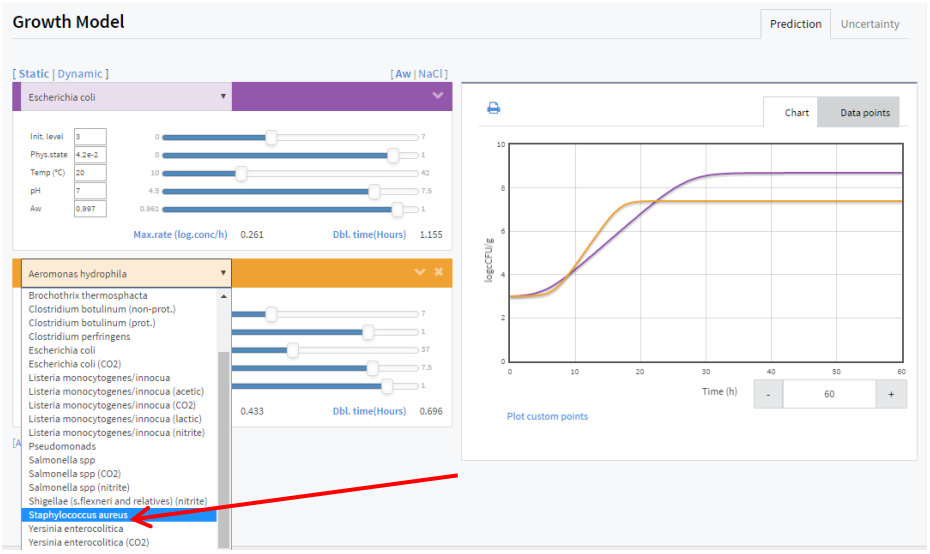


b) Sabendo que temperaturas de 10˚C ou inferiores implicam um enorme consumo de energia, tornando o processo de preservação de alimentos muito dispendioso, explorar outros fatores que possam ser alterados (exemplo o pH e a concentração de cloreto de sódio (NaCl)) de modo a que, mantendo os alimentos à temperatura ambiente (20˚C), as culturas de *Escherichia coli* mantenham um crescimento igual ou inferior ao obtido a 10˚C.

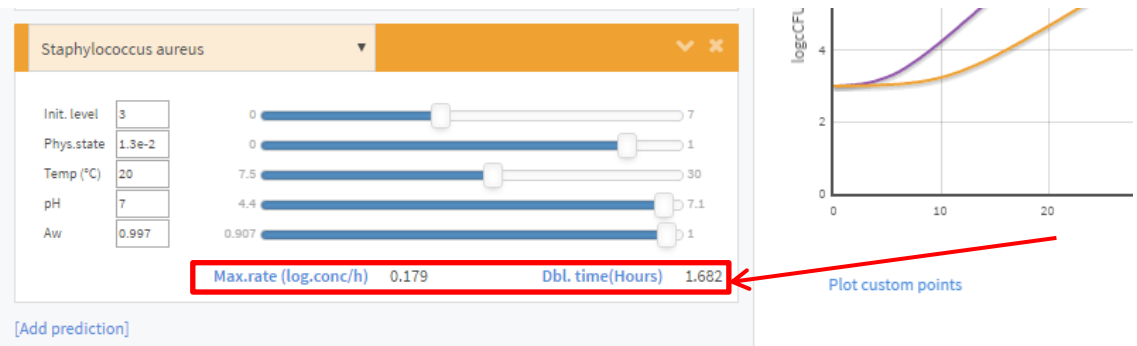
Exercício 18: *ComBase – comparar o crescimento de duas espécies bacterianas*

18.1. Selecione a opção *Add prediction*

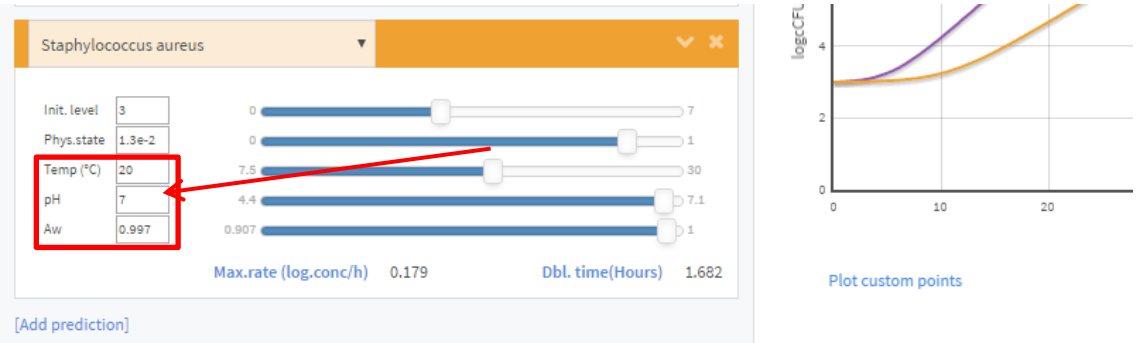
18.2. Escolha a espécie *Staphylococcus aureus*



18.3. Identificar as diferentes fases de crescimento bacteriano e analisar os parâmetros: Max. Rate e Dbl.time



18.4. Alterar as condições em que a *Staphylococcus aureus* se encontra e responder às seguintes a e b, justificando com dados recolhidos na plataforma, nomeadamente o Max. Rate e o Dbl.time:

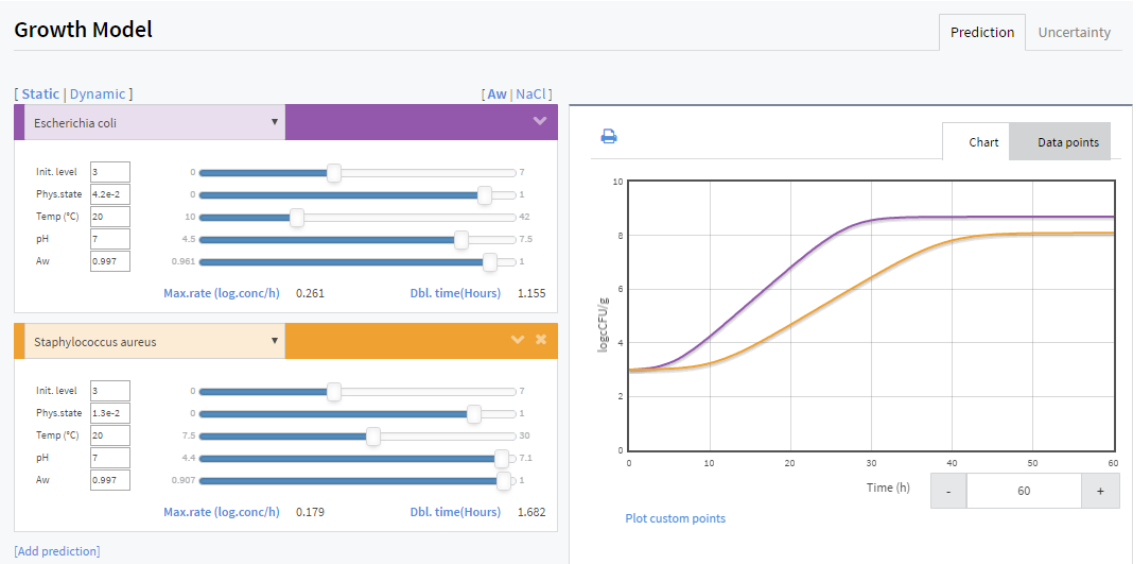


a) Para temperaturas inferiores a 10˚C, o crescimento da bactéria é lento ou rápido?

b) Sabendo que temperaturas de 10˚C ou inferiores implicam um enorme consumo de energia, tornando o processo de preservação de alimentos muito dispendioso, explorar outros fatores que possam ser alterados (exemplo o pH e a concentração de cloreto de sódio (NaCl)) de modo a que, mantendo os alimentos à temperatura ambiente (20˚C), as culturas de *Staphylococcus aureus* mantenham um crescimento igual ou inferior ao obtido a 10˚C.

Exercício 19: *ComBase – modelos de inibição do crescimento bacteriano*

Podendo comparar o crescimento das duas espécies, estude as condições relativas às variáveis: temperatura, pH e Concentração em NaCl, com o objetivo de inibir o crescimento das duas espécies por mais tempo durante um período de 20h.



**Nota sobre a plataforma:*

A plataforma *ComBase* apresenta a vantagem de permitir comparar até 4 modelos de crescimento simultaneamente e a manipulação das variáveis é traduzida no gráfico em tempo real.

**Nota explicativa:

Phys. state ("*initial physiological state*"/estado fisiológico inicial): o valor deste parâmetro pode variar entre 0 e 1 e expressa a adequação física das células ao seu ambiente. Se o seu valor for 0, então o crescimento não ocorrerá (fase *lag* infinita); Se o valor for 1, então o crescimento começará imediatamente (sem fase *lag*).

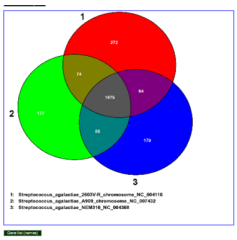
Appendix II: Workshop Guidelines

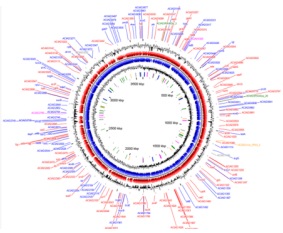
Do DNA aos Genes e à Genómica Comparativa: A Bioinformática na Sala de Aula: Workshop Guidelines

VI Encontro Internacional

Casa das Ciências

Do DNA aos Genes e à Genómica Comparativa: A Bioinformática na Sala de Aula





Fernando Tavares e Ana Sofia Martins

Departamento de Biologia – Faculdade de Ciências da Universidade do Porto

CIBIO - Centro de Investigação em Biodiversidade e Recursos Genéticos/InBIO – Laboratório Associado - UP

VI Encontro Internacional da Casa das Ciências – Lisboa, 10 – 12 de Julho de 2019

O Workshop

Reconhecendo as exigências do programa curricular para a Biologia no ensino secundário, neste workshop procuraremos explorar o potencial da bioinformática como recurso didático para abordar conteúdos como a organização e regulação do material genético, até à temática da evolução. Partindo de uma análise *in silico* de uma sequência de DNA propomos identificar genes e determinar as funções putativas dos seus produtos.

Adicionalmente, recorrendo a recursos bioinformáticos de genómica comparativa será ainda analisada a presença de determinados genes em diferentes grupos taxonómicos de forma a inferir relações evolutivas. Esta atividade contribuirá para uma abordagem holística de noções básicas de genómica, genes e proteínas, assim como elaborar hipóteses evolutivas que possam explicar o mesmo contexto genómico em taxa distintos.

Listagem de Abreviaturas e Siglas

BLAST - Basic Local Alignment Search Tool: esta ferramenta permite, através de um algoritmo matemático, localizar as sequências mais semelhantes à sequência em estudo que se encontram presentes na base de dados.

MaGe - Microbial Genome Annotation & Analysis Platform: https://www.genoscope.cns.fr/agc/microscope/home/index.php

NCBI - National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/

ORF - Open reading frame: sequência de DNA compreendida entre um codão de início (geralmente ATG) da tradução e um codão de terminação, descontando as sequências que correspondem aos intrões no caso de os haver.

EDGAR – *Efficient Database framework for comparative Genome Analyses using BLAST score Ratios*: https://edgar.computational.bio.uni-giessen.de/cgi-bin/edgar_login.cgi?logged=1

BRIG – *BLAST Ring Image Generator*

ANI - *Average Nucleotide Identity*

Biologia Molecular: Análises *in Silico*

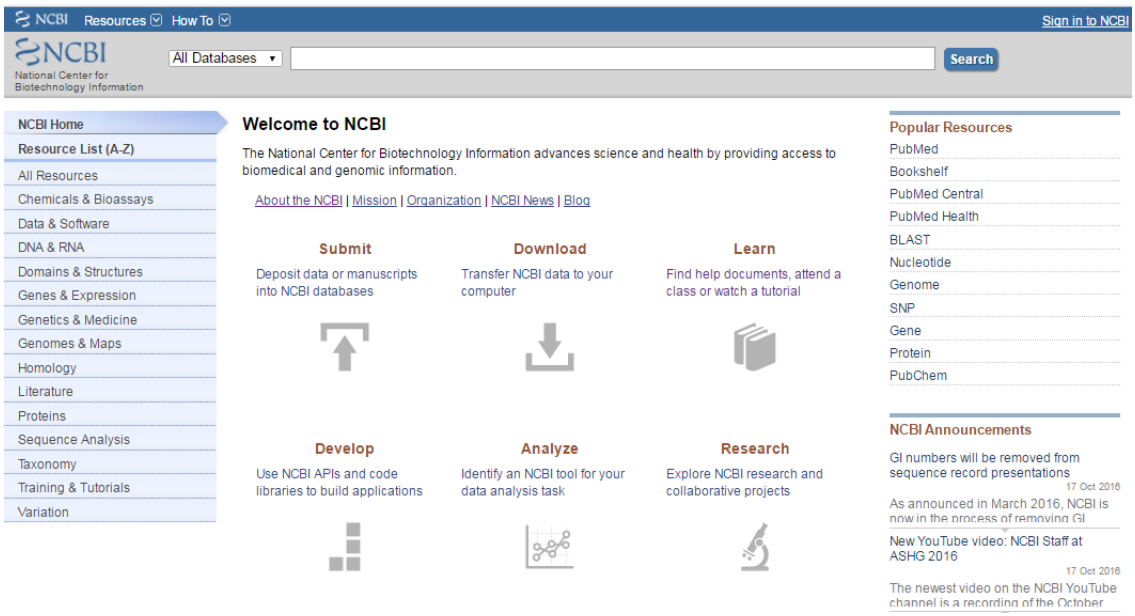
Serão exploradas ferramentas de análise de sequências de DNA. Recursos bioinformáticos, nomeadamente o *NCBI ORF finder* (http://www.ncbi.nlm.nih.gov/orffinder/) e o chamado BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi), serão usados para análise *in silico* de um fragmento de DNA cuja função é desconhecida. Será possível identificar os genes putativos e suas funções, comparando com sequências já conhecidas. Este conhecimento contribuirá para uma abordagem holística de noções básicas como genoma, gene e a importância de regiões intergénicas.

Enquadramento curricular:

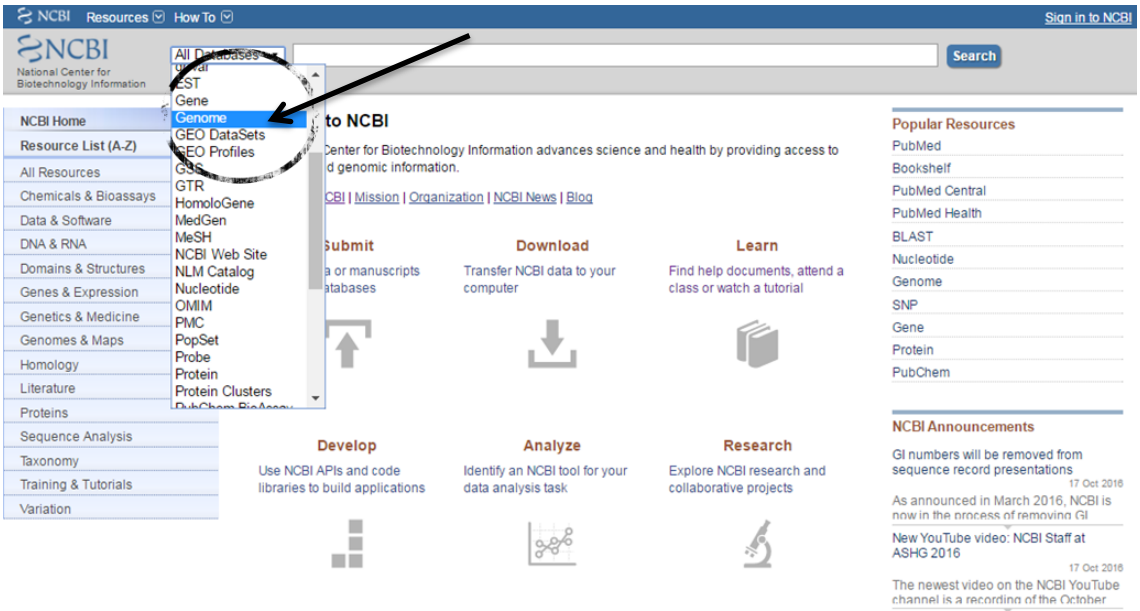
Este exercício permitirá explorar ao nível do 12ºano de escolaridade a organização do material genético, enfatizando, por exemplo, os cromossomas como entidades que contêm genes ou o papel dos operões para os seres procariontes e dos componentes intervenientes em mecanismos de regulação (Módulo 1: Reprodução e Património Genético – 3.2. Organização e regulação do material genético).

Exercício 1: *Aceder a uma sequência de DNA*

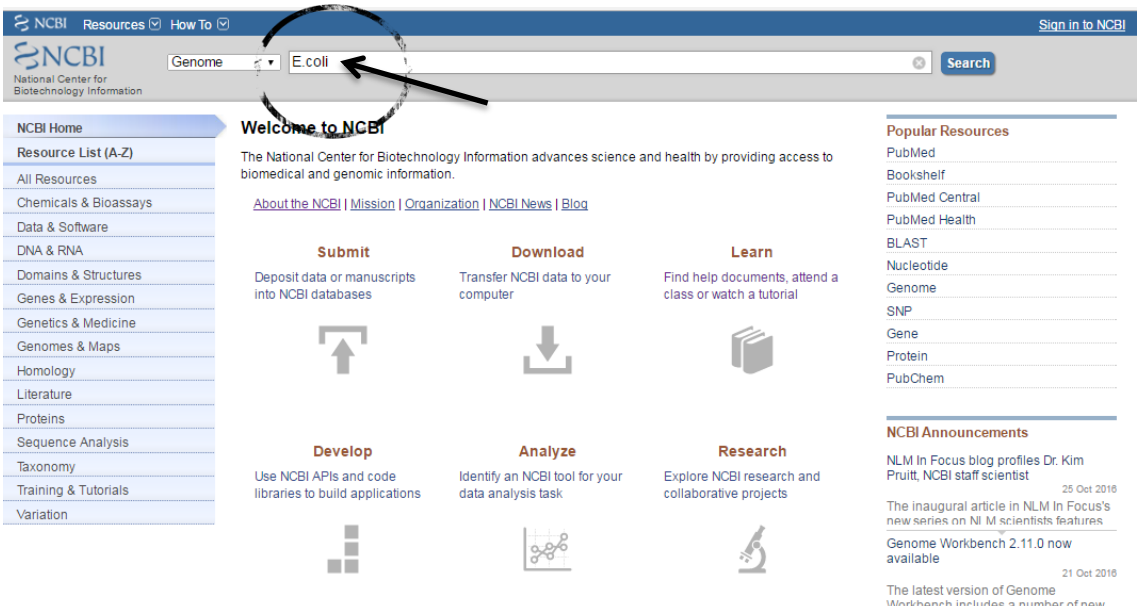
1.1. Aceder à plataforma NCBI através do link: http://www.ncbi.nlm.nih.gov/



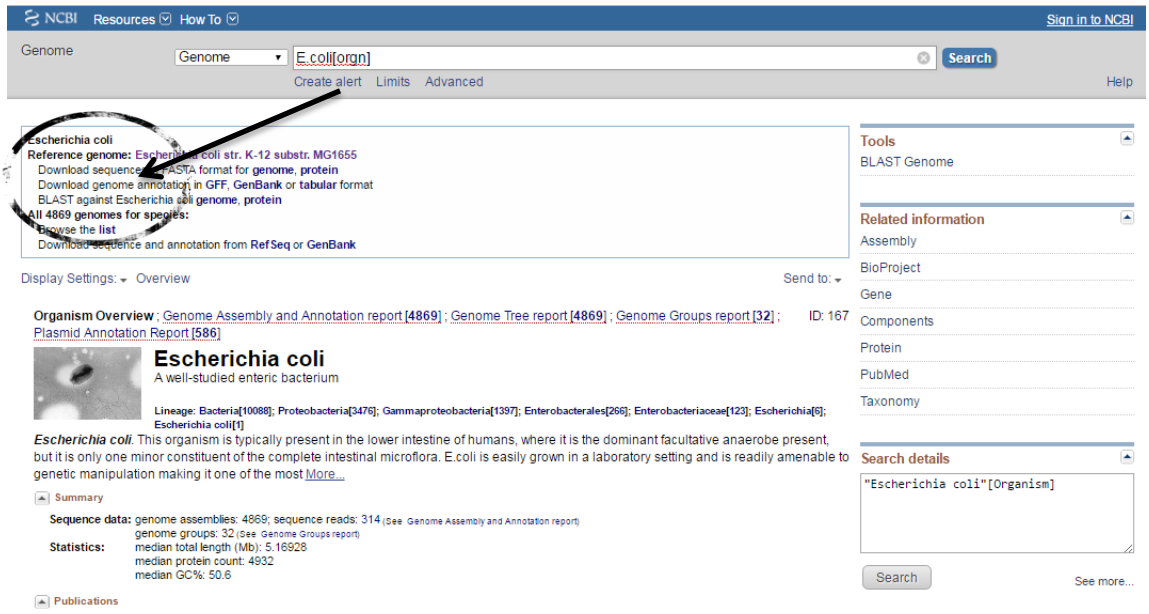
1.2. No menu de busca, selecionar: ***Genome*** (Genoma)



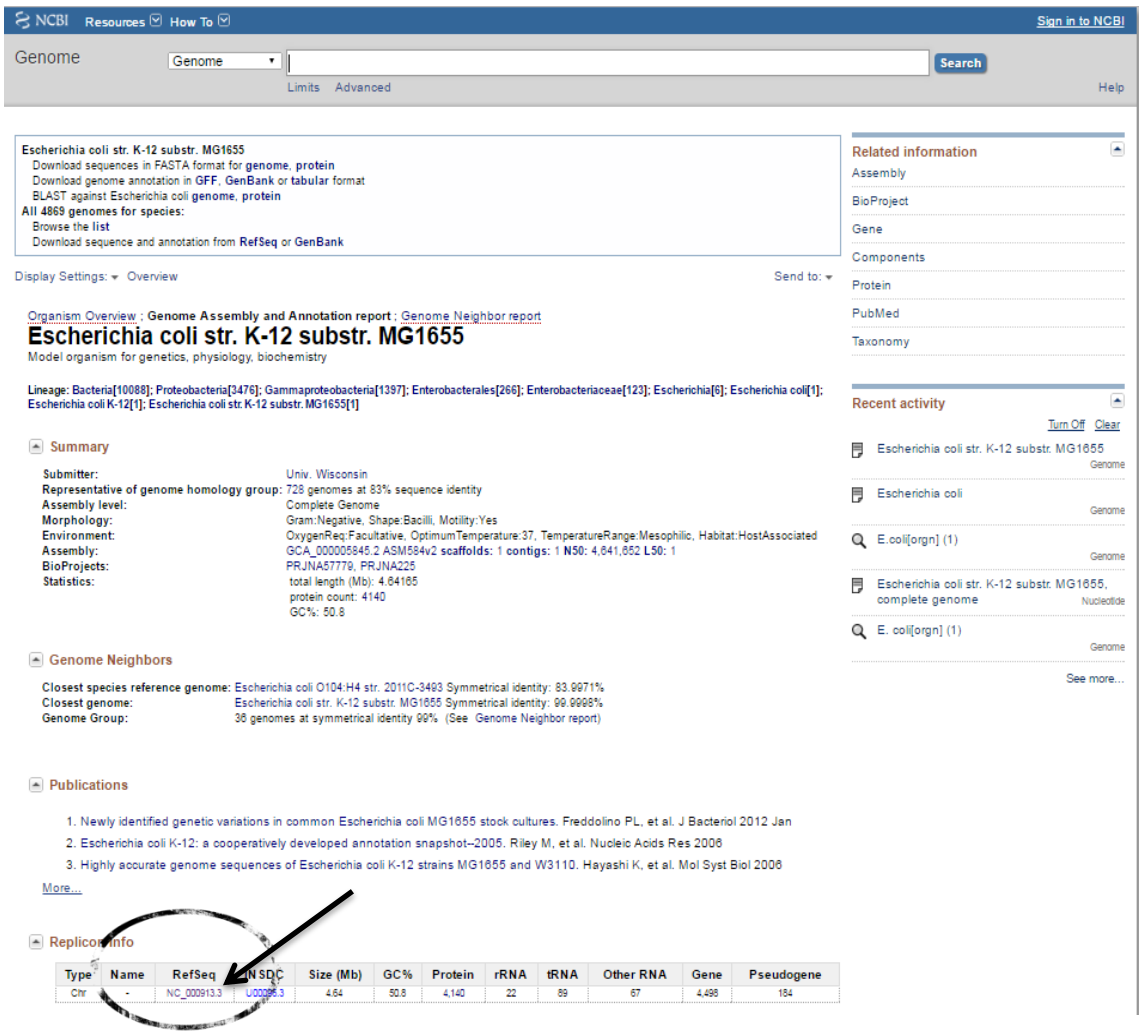
1.3. Procurar por "***E.coli***" e clicar em ***Search*** (pesquisar)



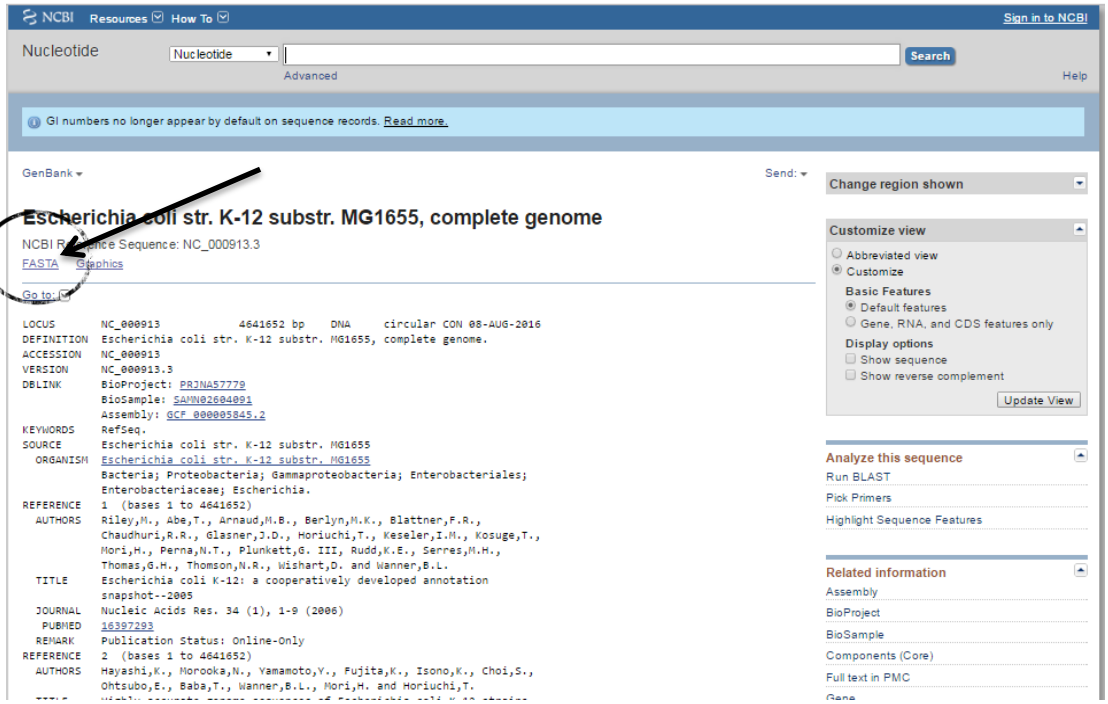
1.4. No início da nova página, selecionar o ***Reference Genome*** (Genoma de Referência)



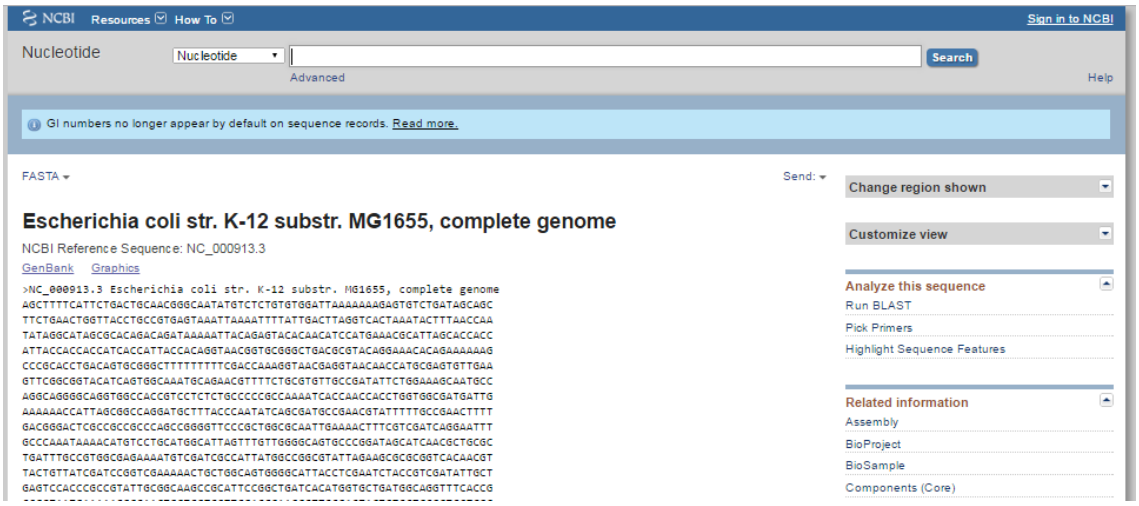
1.5. Descer a página até encontrar o campo: ***Replicon Info*** e clicar em **NC_000913.3**

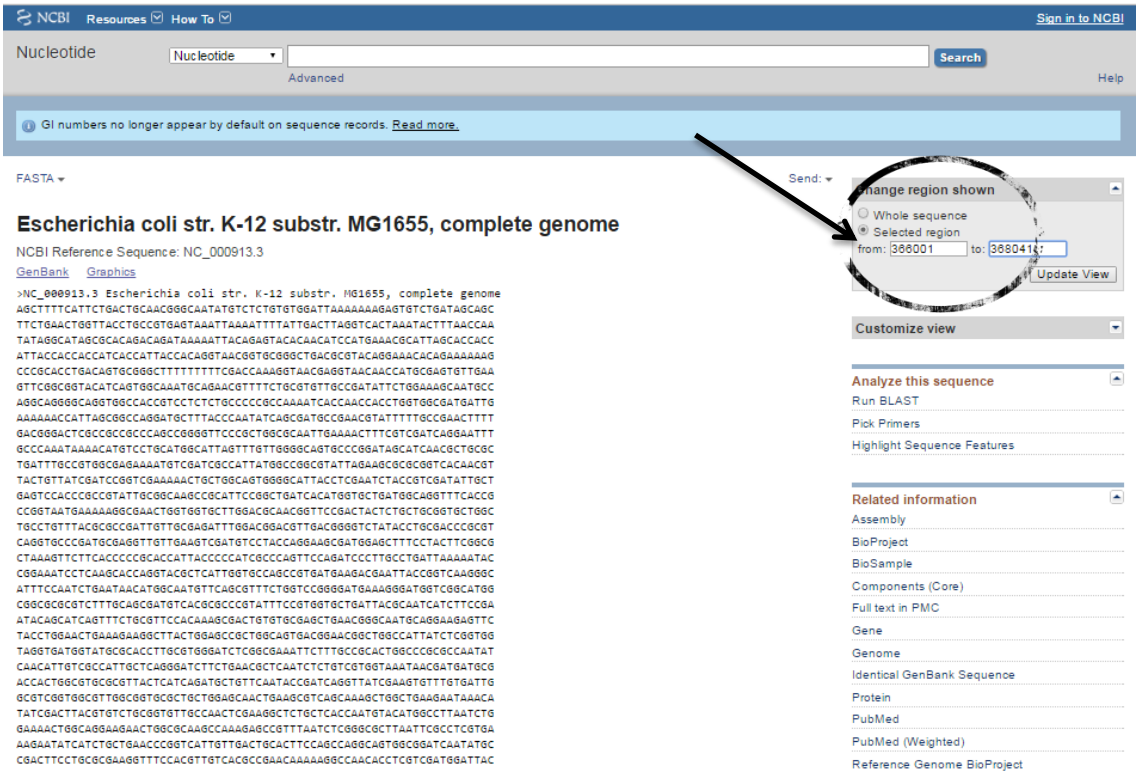


1.6. Escolher a opção de **formato FASTA**

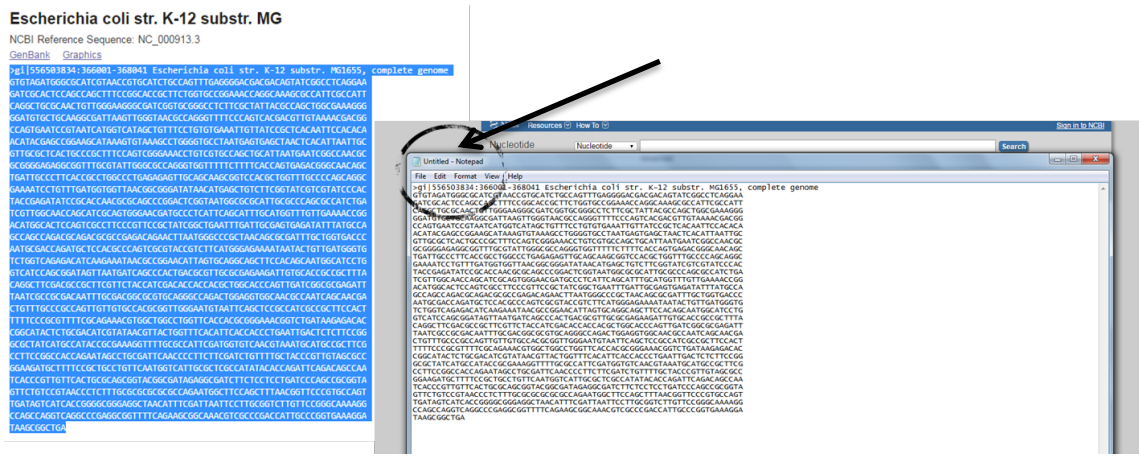


1.7. Abrir a caixa de seleção ***Change region shown*** (alterar região a apresentar) e inserir as coordenadas: **366001-368041**



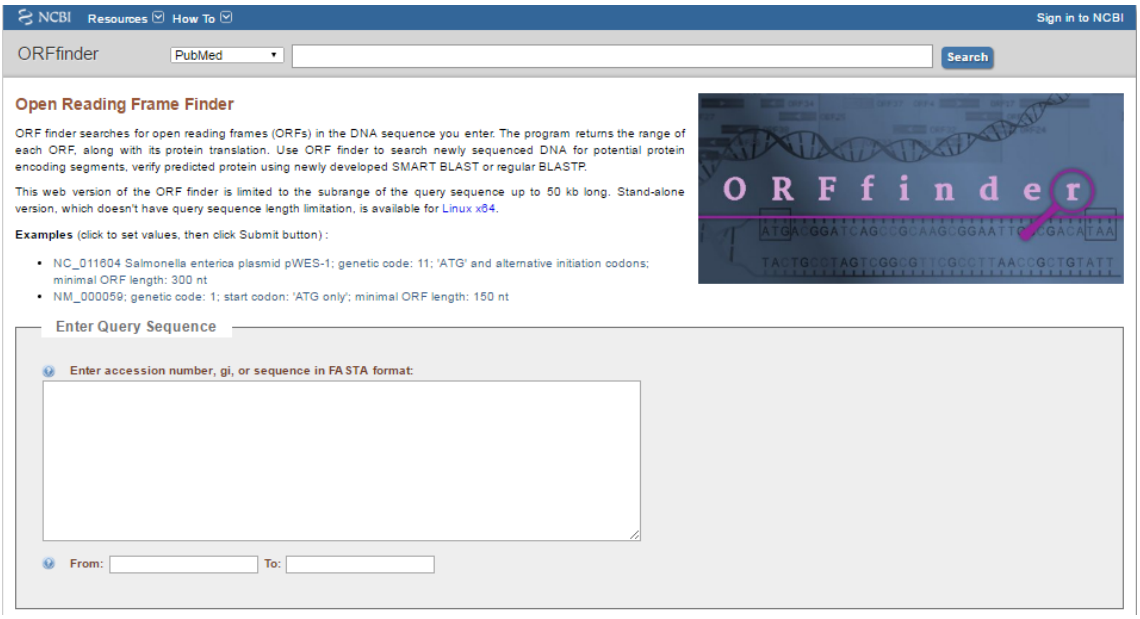


1.8. Copiar e colocar toda a sequência num editor de texto (exemplo: Bloco de Notas)

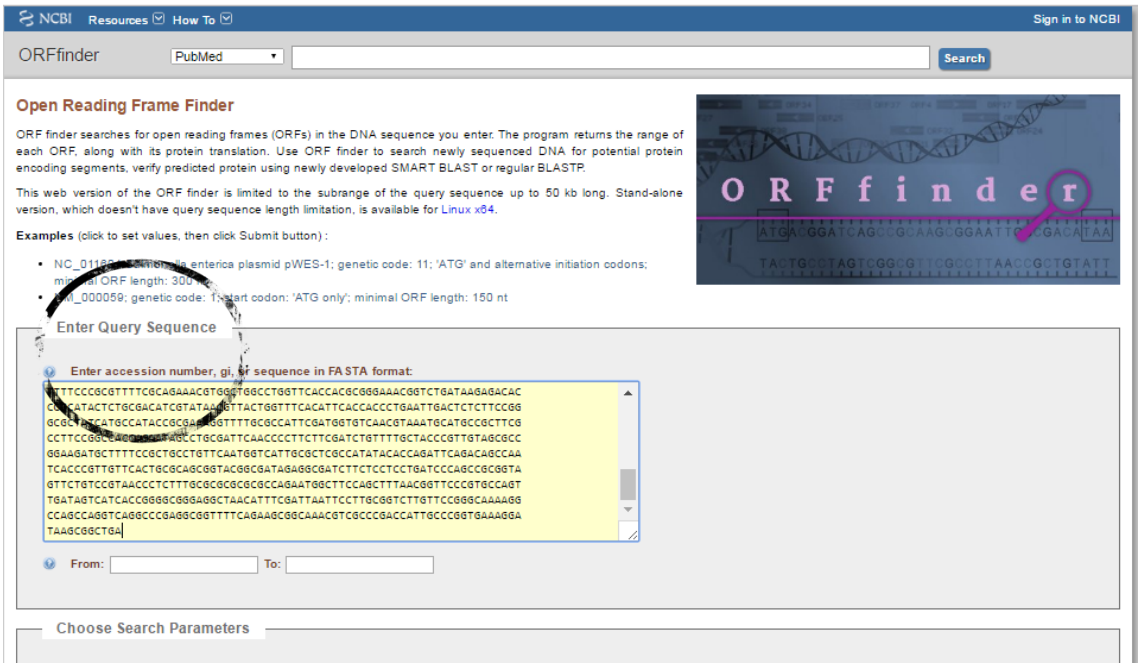


Exercício 2: *Identificação de ORF's (codões de iniciação alternativos)*

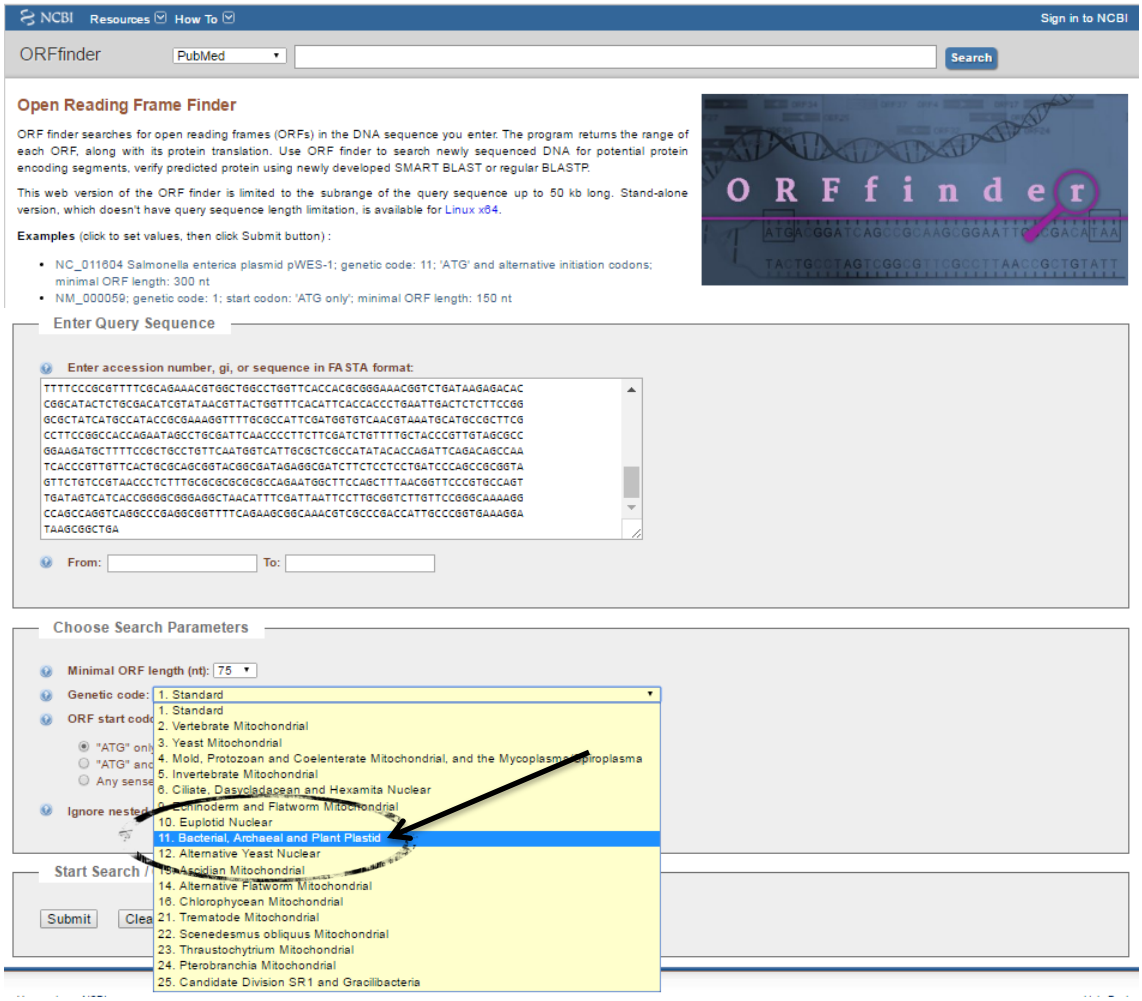
2.1. Aceder à página "*ORFfinder*" da plataforma NCBI através do link: http://www.ncbi.nlm.nih.gov/orffinder/



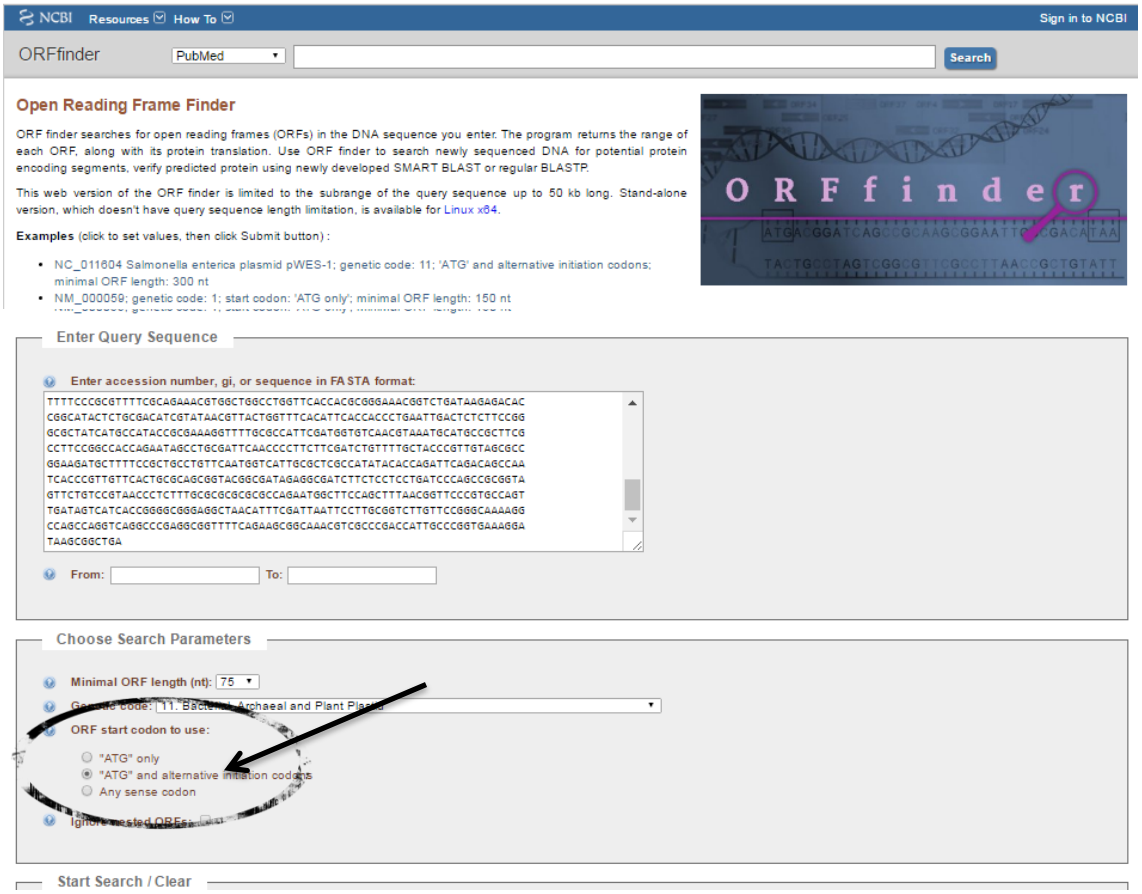
2.2. Inserir a sequência obtida previamente e anotada no processador de texto



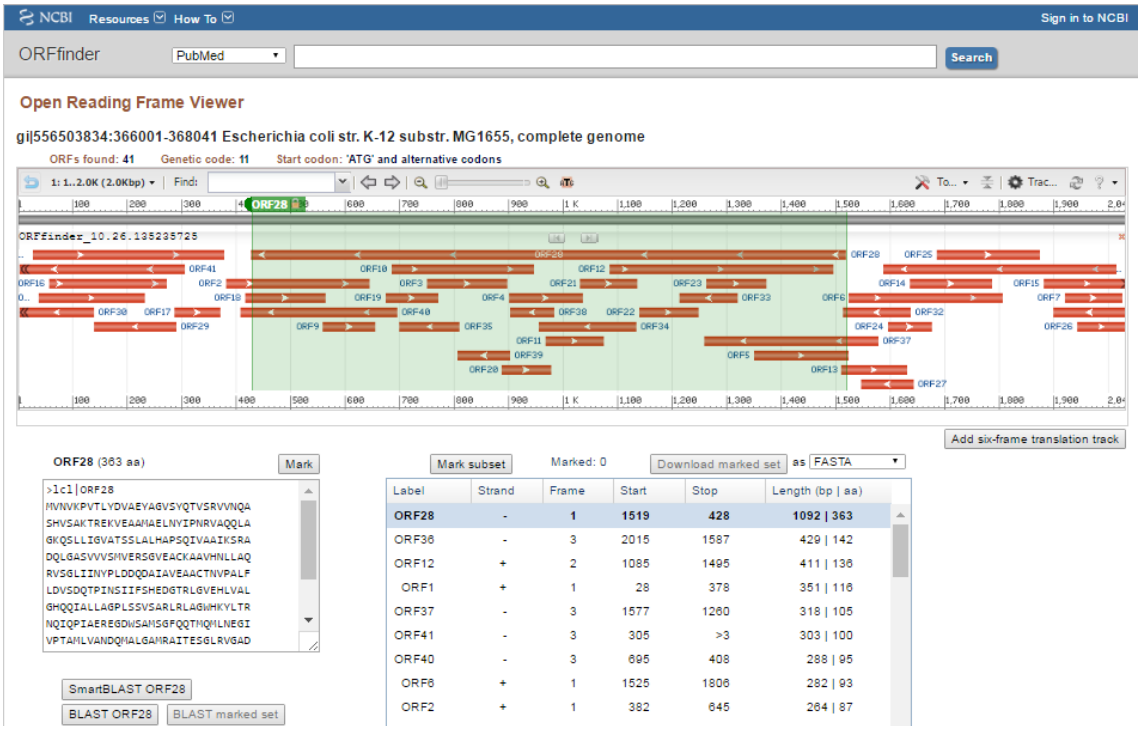
2.3. Escolher o código genético: ***11. Bacterial, Archaeal and Plant Plastid***



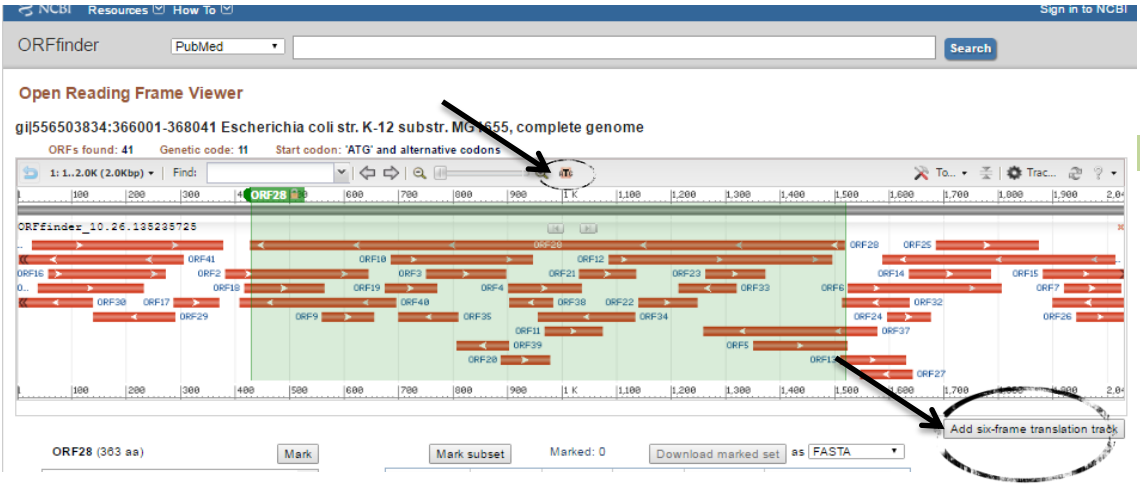
2.4. Escolher a opção: ***"ATG" and alternative initiation codons*** ("ATG" e codões de iniciação alternativos) e clicar em ***Submit*** (Submeter)



2.5. Analisar os resultados obtidos



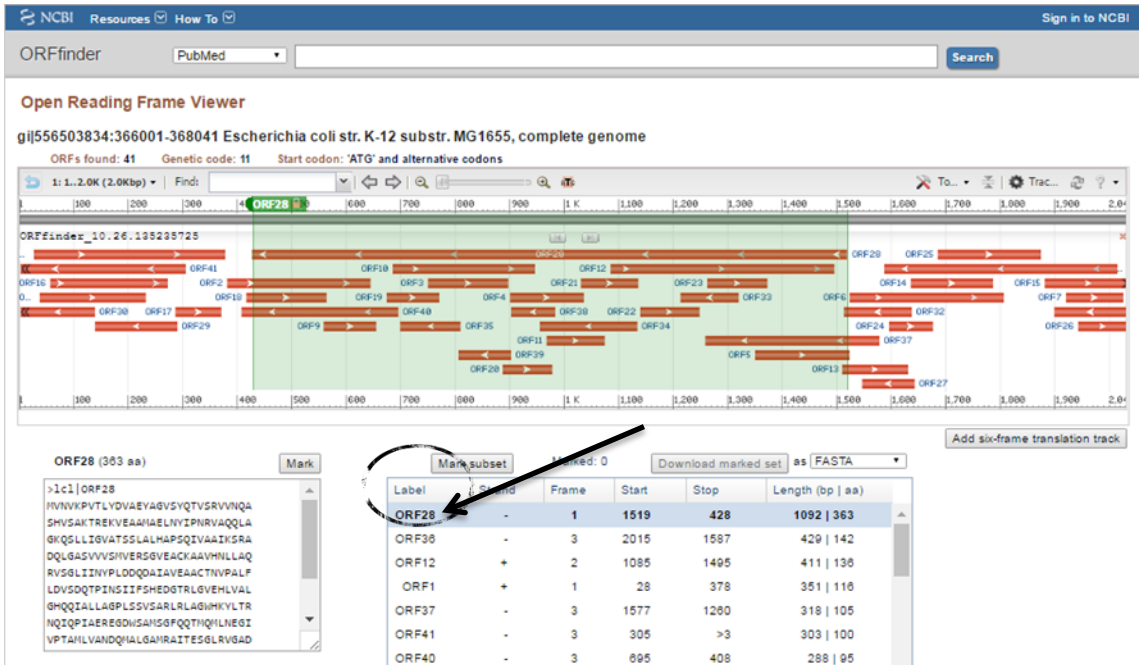
2.6. Abrir a opção ***Add six-frame translation track*** e explorar a função ***ATG*** (não fechar a página e prosseguir para o exercício 3)



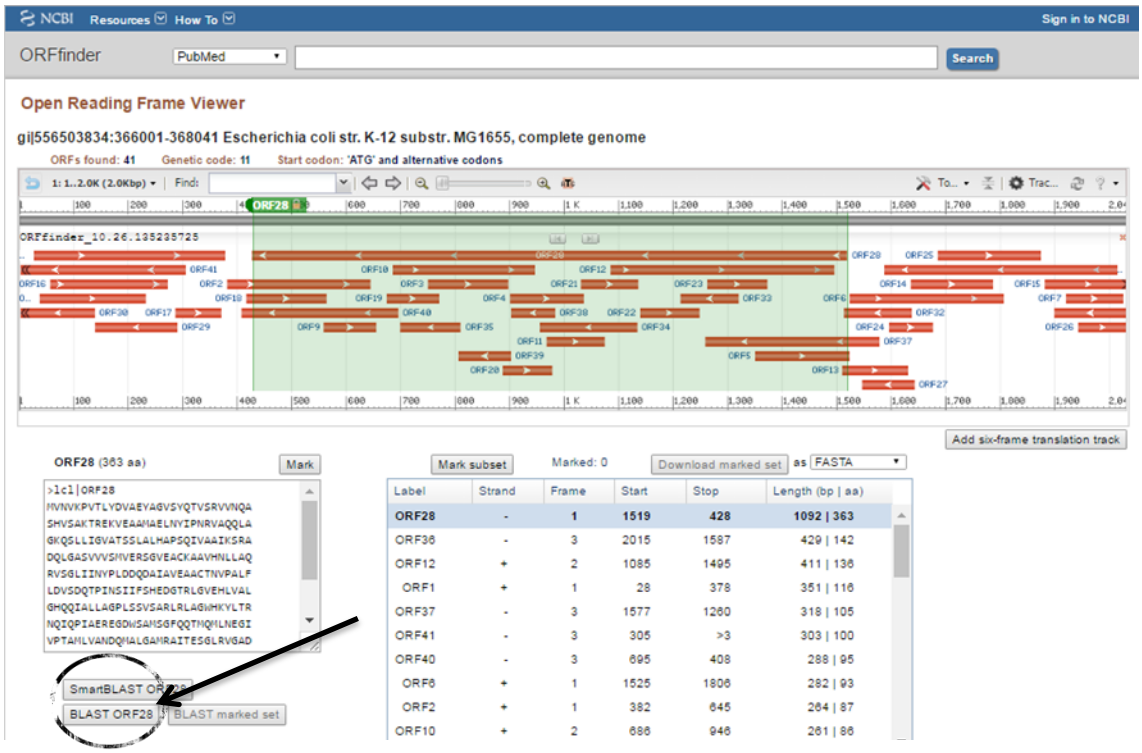
Exercício 3: *Das ORF's a genes putativos*

3.1. Analisar as características das ORF's obtidas.

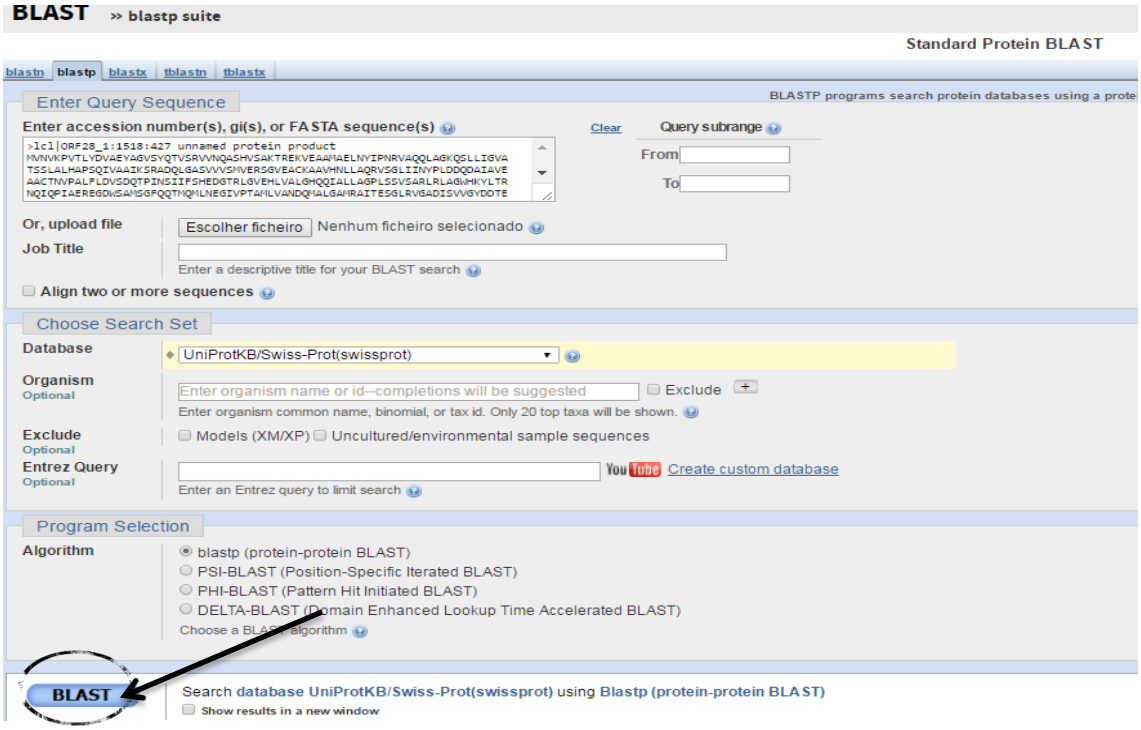
3.2. Selecionar uma ORF clicando no repetitivo código (exemplo: ORF28)



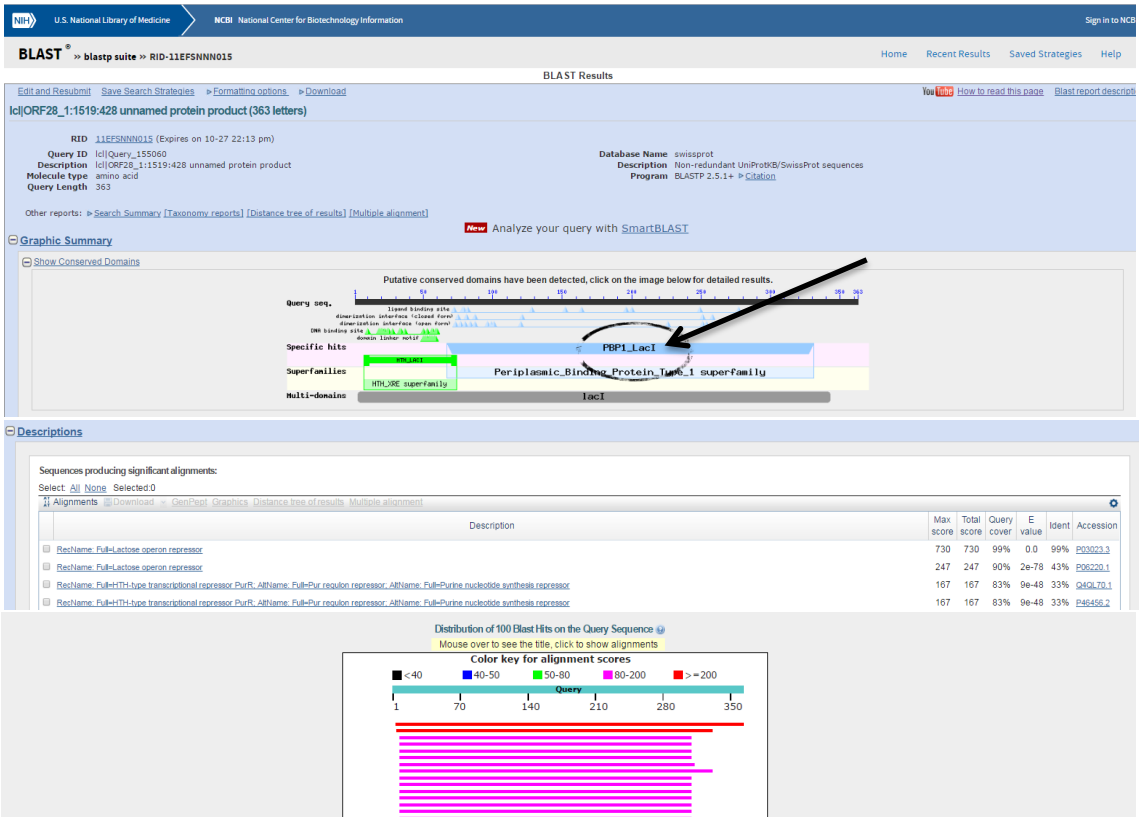
3.3. Iniciar o BLAST da ORF selecionada clicando em ***BLAST ORF*** (será redirecionado para uma página do *NCBI BLAST tool*: https://blast.ncbi.nlm.nih.gov/Blast.cgi)



4.4. Na página aberta clicar em ***BLAST*** para iniciar



3.5. Identificar o gene.



3.6. Repetir os passos 3.2. a 3.5. para duas ORF's alternativas.

Operão *lac*: Regulação Génica e Relações Evolutivas

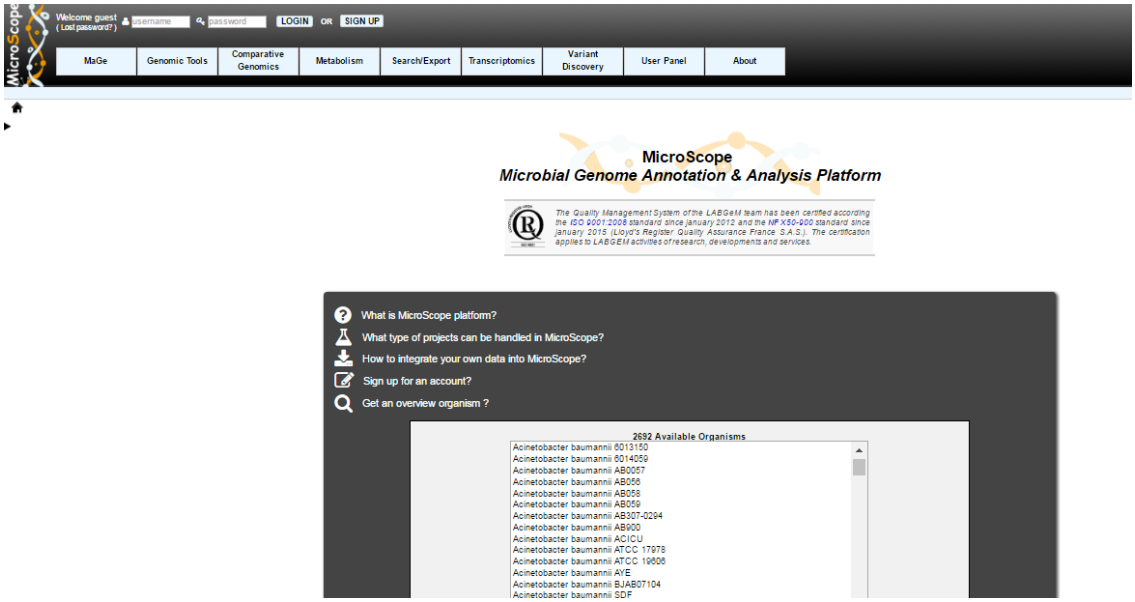
Utilizando uma plataforma bioinformática de genómica comparativa (MaGe – MicroScope - https://www.genoscope.cns.fr/agc/microscope/home/index.php) ser possível investigar relações evolutivas, identificando a presença dos genes do operão lac e das regiões flanqueantes em diferentes grupos taxonómicos. Este exercício permitirá elaborar hipóteses evolutivas que possam explicar a presença dos genes em taxa distintos.

Enquadramento curricular:

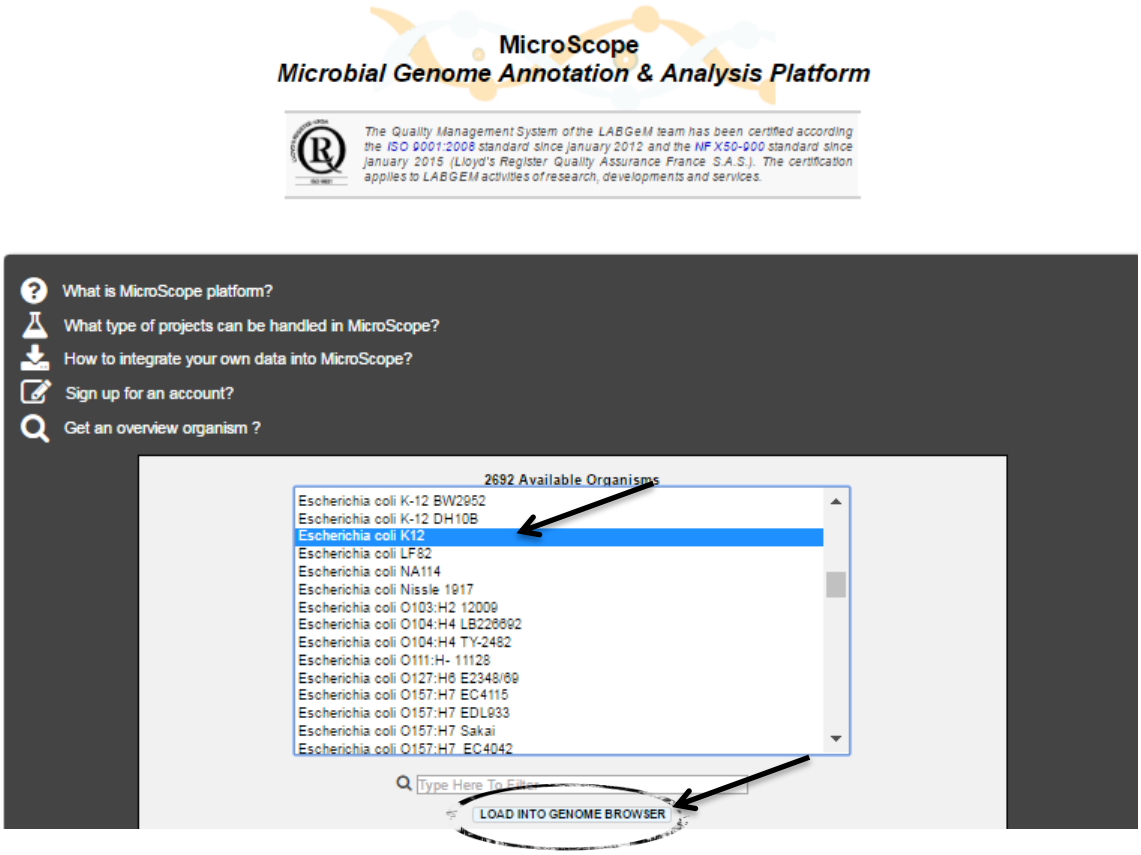
Será possível abordar no 11ºano de escolaridade a Unidade 7: Evolução Biológica.

Exercício 4: *Genómica Comparativa*

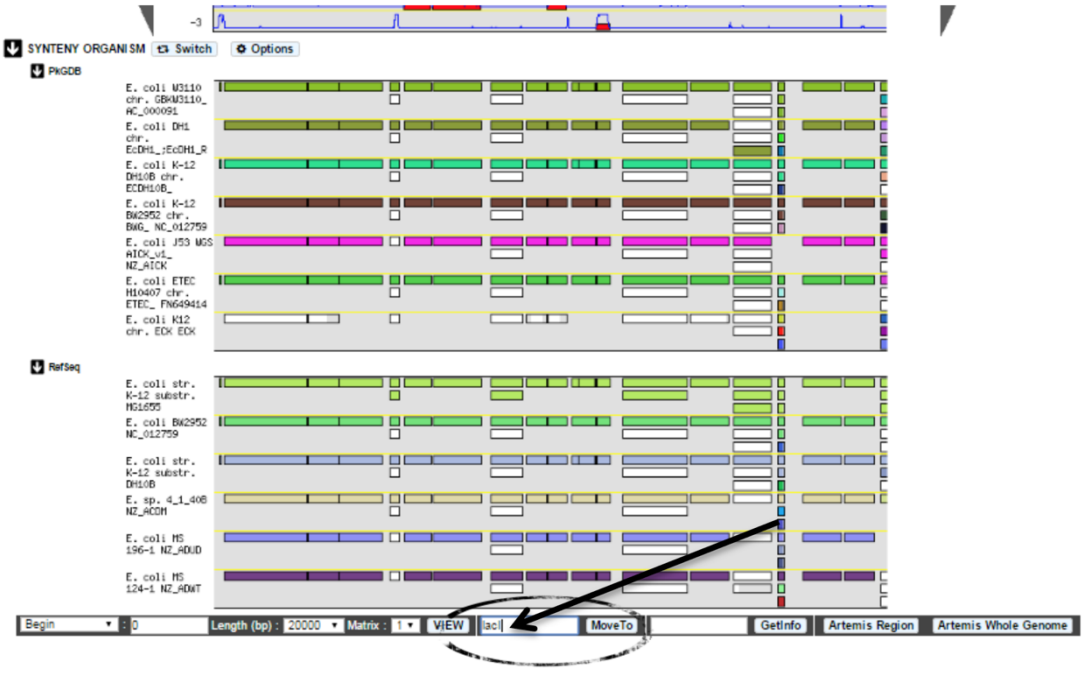
4.1. Aceder à plataforma MicroScope através do link: https://www.genoscope.cns.fr/agc/microscope/home/index.php



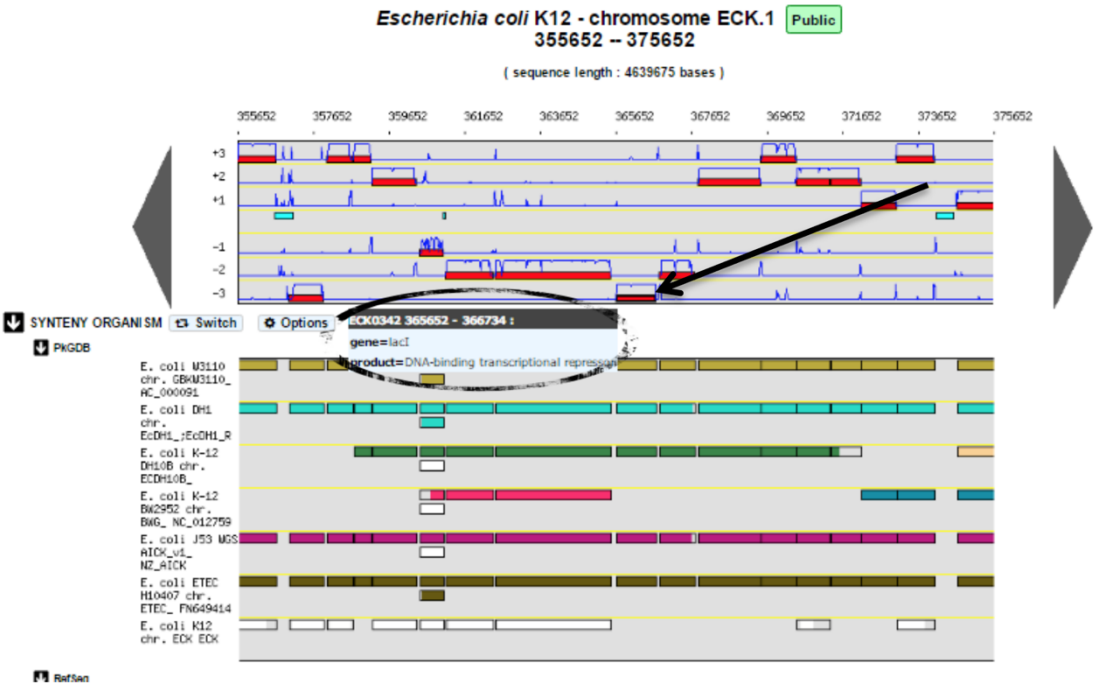
4.2. Escolher: ***Escherichia coli K12*** e clicar em ***load into genome browser*** (carregar no motor de busca de genomas)



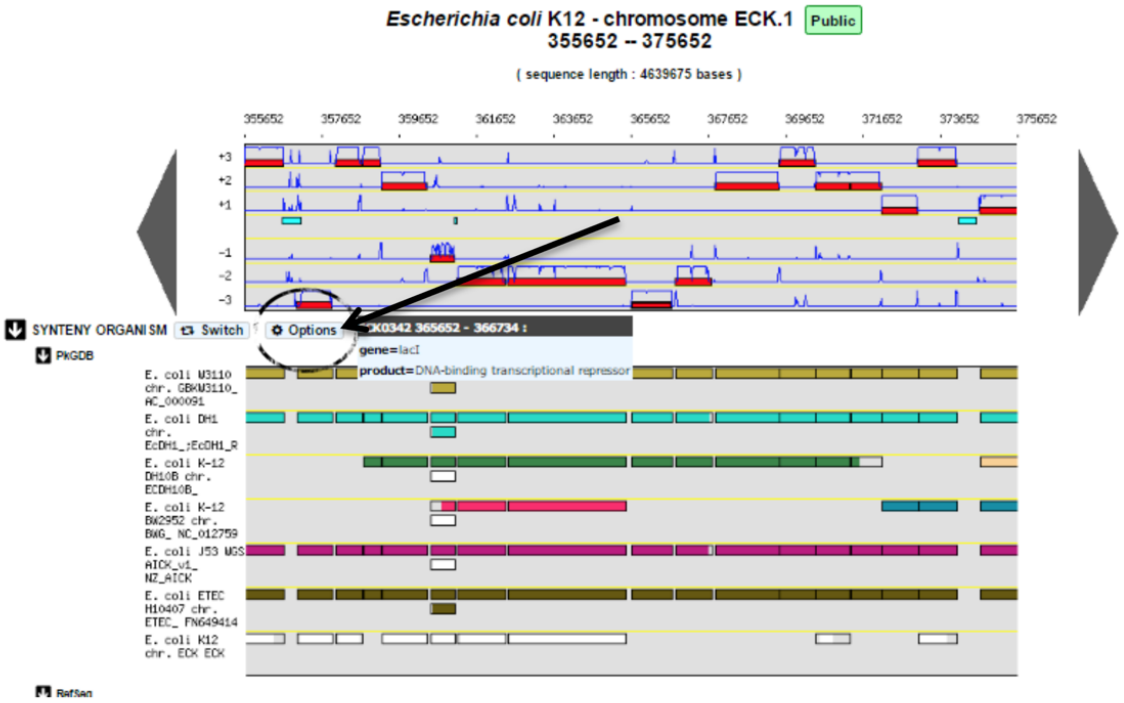
4.3. Para localizar o gene, pesquisar por ***lacI*** e clicar em: ***Move to*** (Mover para) na barra abaixo ilustrada.



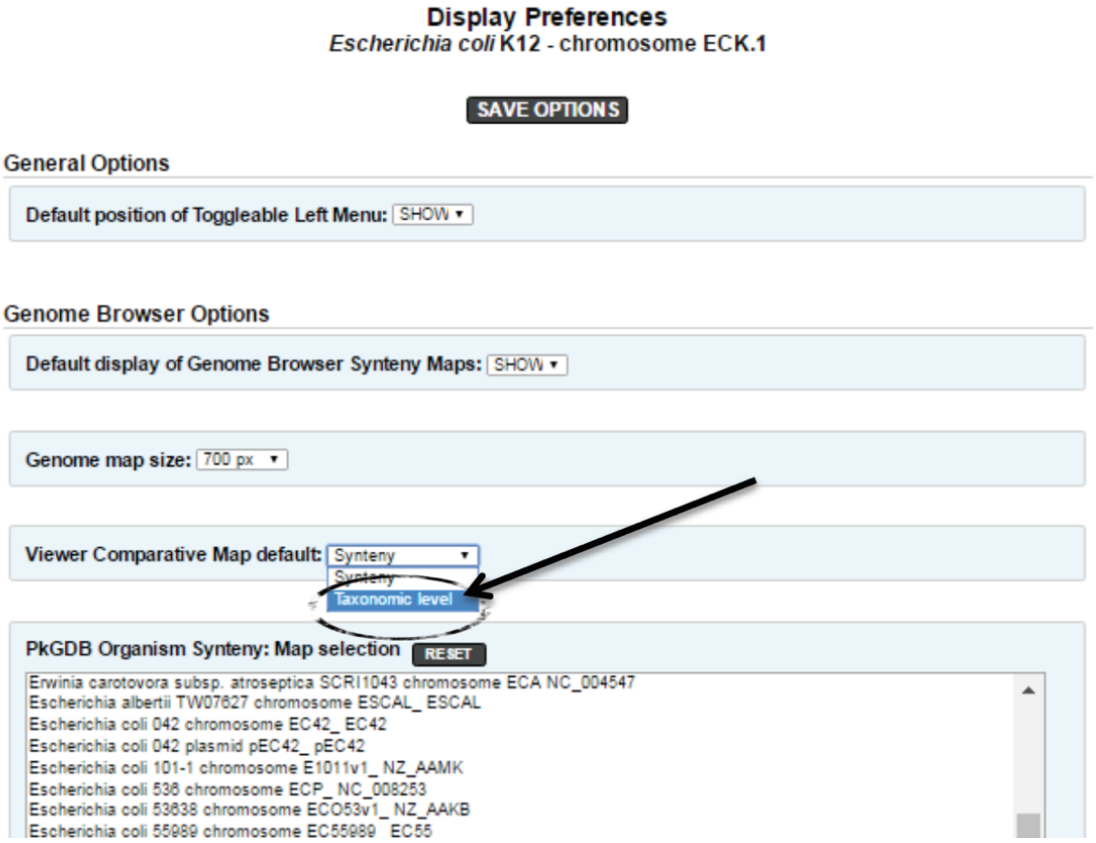
4.4. Identificar o gene *lacI* movendo o rato por cima das barras vermelhas



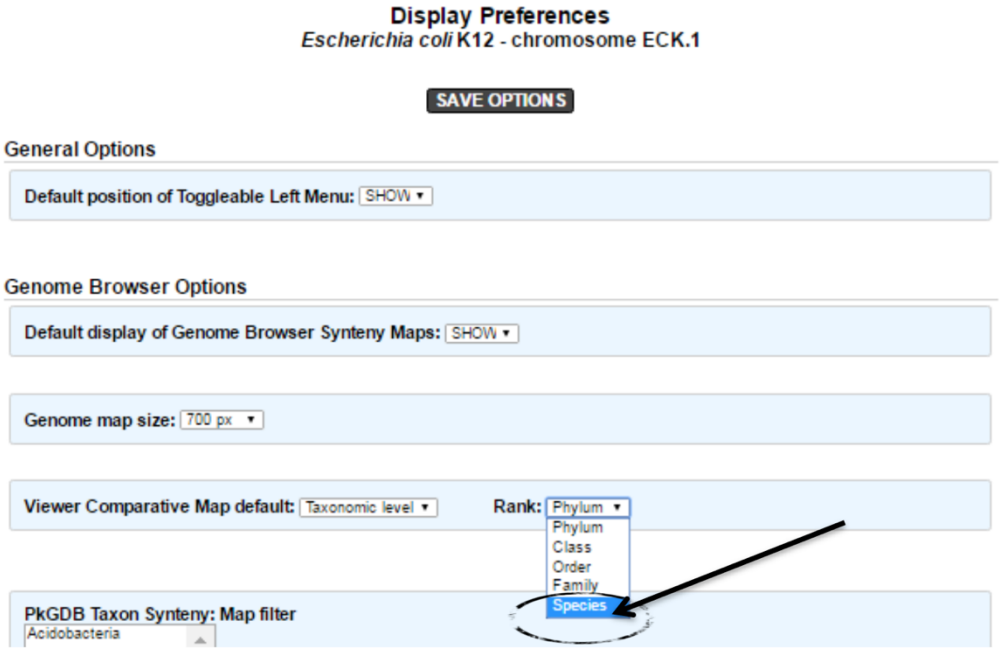
4.5. Selecionar ***Options Menu*** (Menu de opções)



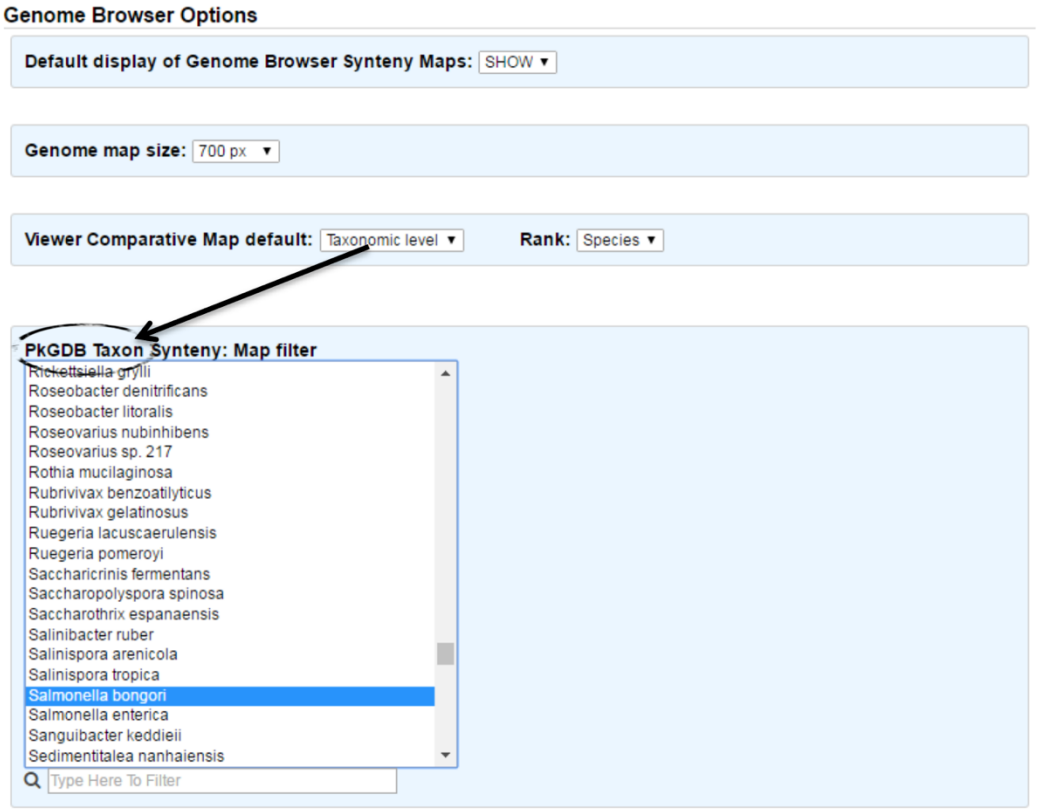
4.6. Na nova janela aberta, na secção ***Viewer Comparative Map default***, escolher ***taxonomy level*** (nível taxonómico)



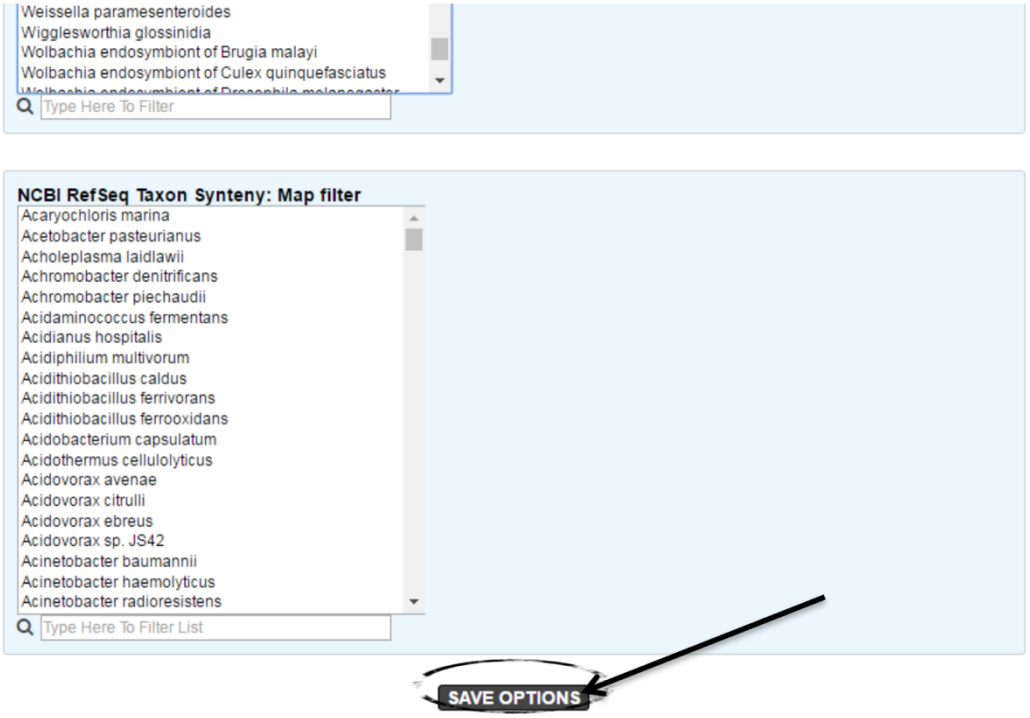
4.7. Em ***Rank*** (Ordenar), escolher ***Species*** (Espécies)



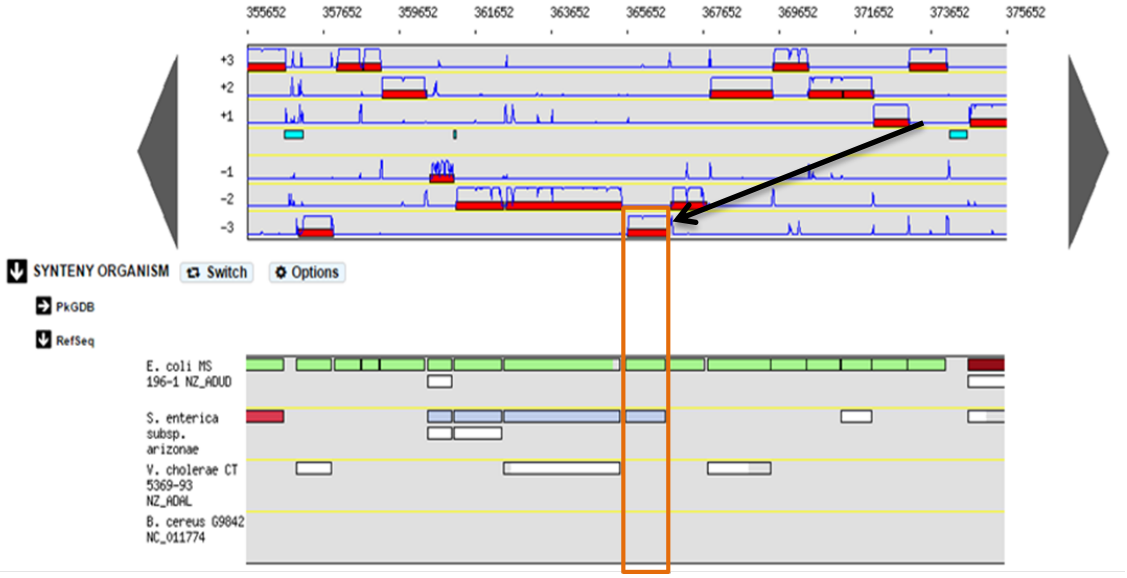
4.8. Na secção ***PkGDB Taxon Synteny: Map filter***, pressionar o botão CTRL e escolher *Bacillus cereus, Escherichia sp.4_1_40B, Samonella bongori e Vibrio fischeri.*



4.9. Clicar em ***Save options*** (Guardar opções)



4.10. Comparar a presença e a função do gene em diferentes *taxa*.



Importância dos Genes e das Mutações no Estudo da Evolução

Propõe-se a utilização de ferramentas bioinformáticas e bases de dados disponibilizados no EDGAR (https://edgar.computational.bio.uni-giessen.de/). Escolhendo até cinco estirpes bacterianas pretende-se identificar o conjunto de genes homólogos e genes específicos de cada estirpe utilizando a funcionalidade Diagramas de Venn (*Venn Diagram*) na plataforma EDGAR. Com base nos resultados obtidos discutir as noções de genoma centro (core genoma), pan genoma e genoma acessório. Será ainda analisado um Diagrama Circular (*Circular Plot*) e a matriz de Média de Identidade de Nucleótidos (*ANI – Average Nucleotide Identity - Matrix*) numa perspetiva de genómica comparativa. Por último, será explorada uma funcionalidade do EDGAR para a criação de árvores filogenéticas.

Enquadramento curricular:

Esta proposta enquadra-se particularmente no 11ºano de escolaridade na Unidade 5: Crescimento e renovação celular – 1.1. DNA e Síntese Proteica; Unidade 7: Evolução Biológica; Unidade 8: Sistemática dos Seres Vivos - 1.2. Taxonomia e Nomenclatura.

Informação sobre as estirpes a estudar:

No primeiro exercício foi pesquisado o genoma da *Escherichia coli* no NCBI (https://www.ncbi.nlm.nih.gov/genome/?term=escherichia+coli).

Na descrição desta espécie bacteriana é possível ler que:

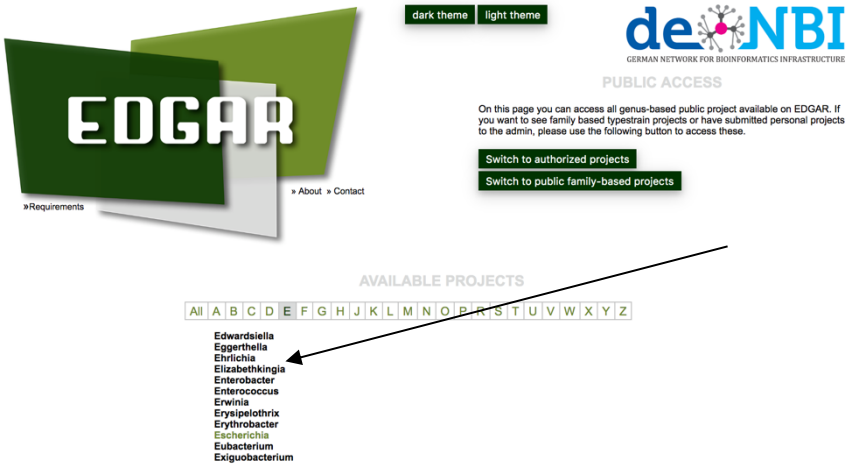
*"**Escherichia coli.*** Este organismo está tipicamente presente no intestino delgado dos seres humanos, onde é o anaeróbio facultativo dominante presente, mas é apenas um constituinte secundário da microflora intestinal completa. A *E. coli* é cultivada em laboratório e é facilmente manipulável geneticamente, tornando-a um dos organismos modelo procariotas mais estudados. *E.coli* K-12 é a estirpe mais amplamente estudada de *E. coli* e serve como referência para esta espécie. *E. coli* é também uma das espécies microbianas mais diversas, contendo estirpes patogénicas e não patogénicas. A estirpes patogénicas podem causar infeções do trato urinário, meningite neonatal e várias doenças intestinais, frequentemente graves. As estirpes patogénicas geralmente apresentam fatores de virulência codificados em plasmídeos, em profagos ou em regiões genómicas denominadas ilhas de patogenicidade (PAIs). É provável que as PAIs tenham resultado de transferência horizontal e possam ter sido integradas no cromossoma pela ação de elementos genéticos móveis. Entre as estirpes patogénicas de *E. coli* podem ser mencionadas: a *Escherichia coli* UMN026 frequentemente associada a infeções do trato urinário, a *Escherichia coli* O157: H7 str. Sakai que causa colite hemorrágica, a *Escherichia coli* O83: H1 str. NRG 857C que está relacionada com a Doença de Crohn e *Escherichia coli* O104: H4 str. 2011C-3493 que causa síndrome hemolítica urémica." *(tradução adaptada)*

Exercício 1: *Venn Diagram*

1.1. Aceder ao site da plataforma EDGAR: https://edgar.computational.bio.uni-giessen.de/

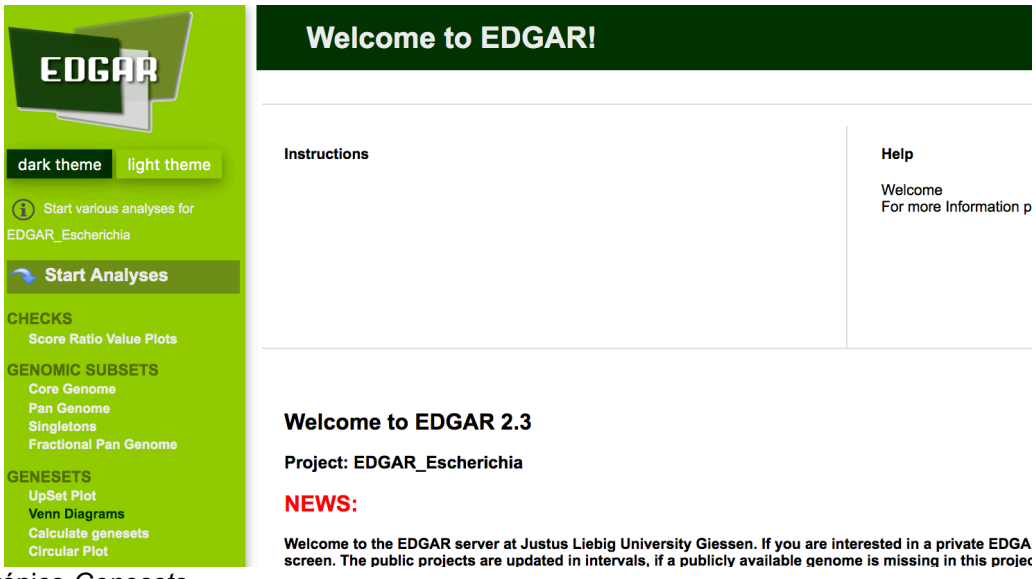
Poderá também ser acedido através do motor de buscar Google digitando: EDGAR Bioinformatics

1.2. Selecionar a letra **E**



1.3. Selecionar o Género ***Escherichia***

1.4. Selecionar na barra verde do lado esquerdo da página a opção ***Venn Diagrams*** no



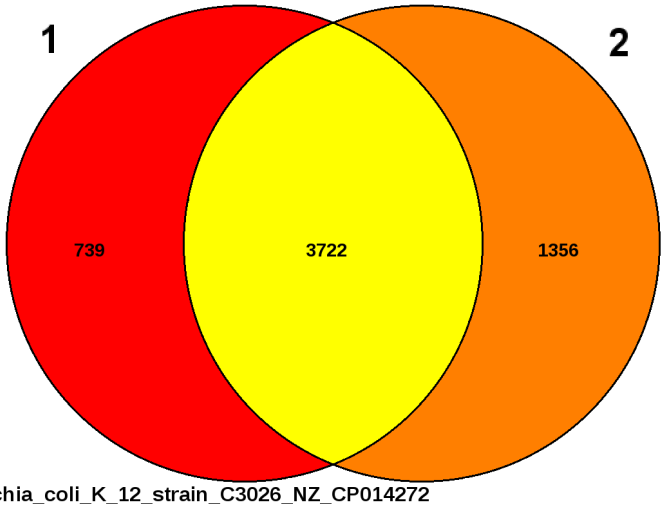
tópico *Genesets*.

1.5. Selecionar as estirpes:

Escherichia_coli_K_12_strain_C3026_NZ_CP014272 *(Não patogénica)*

Escherichia_coli_O157_H7_str_Sakai_NC_002695 (P*atogénica)*

1.6. Analisar os resultados e explorar as noções de ***Core genome*** e ***Pan genome***



1: Escherichia_coli_K_12_strain_C3026_NZ_CP014272
2: Escherichia_coli_O157_H7_str_Sakai_NC_002695

1.7. Voltar ao menu de seleção de estripes e selecionar:

Escherichia_coli_K_12_strain_C3026_NZ_CP014272 *(Não patogénica)*

Escherichia_coli_O157_H7_str_Sakai_NC_002695

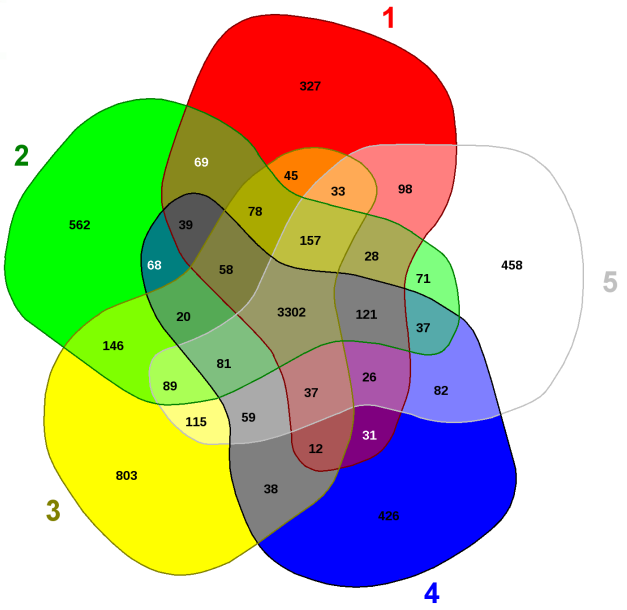
Escherichia_coli_UMN026_NC_011751

Escherichia_coli_O83_H1_str_NRG_857C_NC_017634

Escherichia_coli_O104_H4_str_2011C_3493_NC_018658

Patogénica

1.8. Analisar os resultados. Identificar os genes comuns apenas às bactérias patogénicas. Clicar no número de genes comuns e estudar os genes anotados. Estão relacionados com a processos de infeção?



1: Escherichia_coli_K_12_strain_C3026_NZ_CP014272
2: Escherichia_coli_O104_H4_str_2011C_3493_NC_018658
3: Escherichia_coli_O157_H7_str_Sakai_NC_002695
4: Escherichia_coli_O83_H1_str_NRG_857C_NC_017634
5: Escherichia_coli_UMN026_NC_011751

Comparative set

81 genes found in the genomes of

Escherichia_coli_O104_H4_str_2011C_3493_NC_018658
Escherichia_coli_O157_H7_str_Sakai_NC_002695
Escherichia_coli_O83_H1_str_NRG_857C_NC_017634
Escherichia_coli_UMN026_NC_011751

but not in

Escherichia_coli_K_12_strain_C3026_NZ_CP014272

EXPORT GENE LIST (NAMES) | EXPORT GENE LIST (DNA FASTA) | EXPORT GENE LIST (AA FASTA)

Search:

Escherichia_coli_O104_H4_str_2011C_3493_NC_018658	Escherichia_coli_O157_H7_str_Sakai_NC_002695	Escherichia_coli_O83_H1_str_NRG_857C_NC_017634	Escherichia_coli_UMN026_NC_011751
O3K_16830 formate transporter	ECs0987 formate transporter	NRG857_04115 formate transporter	ECUMN_1097 formate transporter
O3K_16835 formate acetyltransferase	ECs0986 formate acetyltransferase 1	NRG857_04110 pyruvate formate lyase I	ECUMN_1096 pyruvate formate lyase I
O3K_16910 thioredoxin reductase	ECs0973 thioredoxin reductase	NRG857_04050 thioredoxin reductase	ECUMN_1083 thioredoxin reductase, FAD/NAD(P)-binding
O3K_17525 Host specificity protein J of prophage	ECs0842 host specificity protein	NRG857_07680 putative host specificity protein	ECUMN_0631 host specificity protein J

Exercício 2: *Circular Plot (BRIG - BLAST Ring Image Generator)*

2.1. Selecionar na barra verde do lado esquerdo da página a opção ***Circular Plot*** no tópico *Genesets*

2.2. Selecionar as estirpes arrastando para a janela do lado direito:

Escherichia_coli_K_12_strain_C3026_NZ_CP014272 *(Não patogénica)*

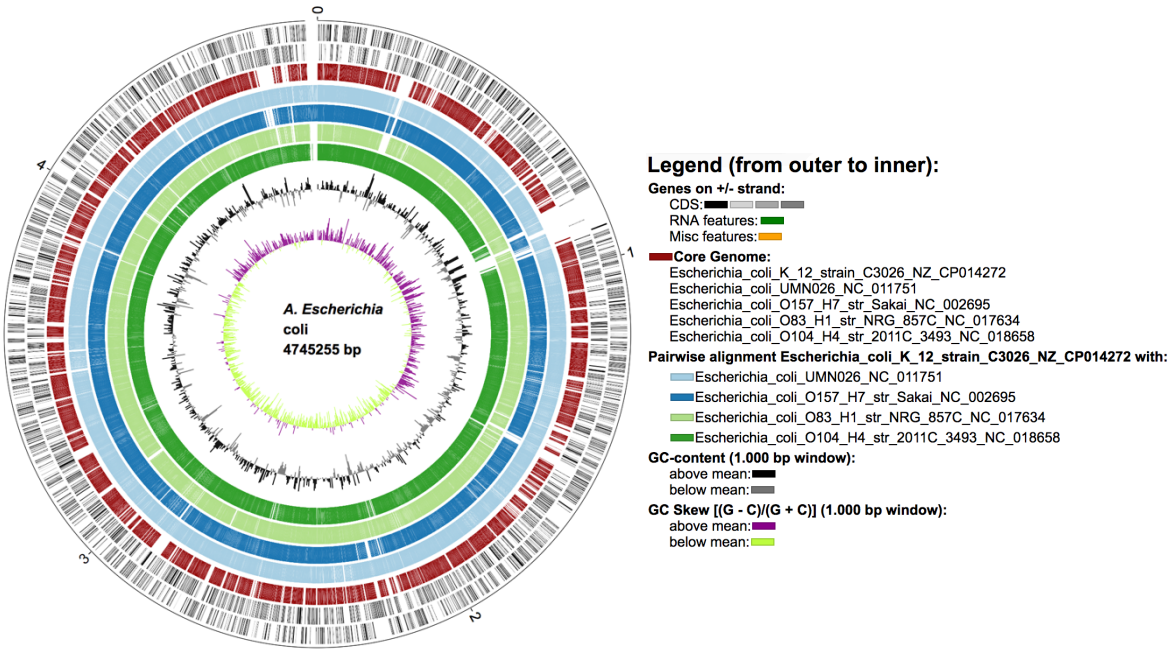
Escherichia_coli_O157_H7_str_Sakai_NC_002695

Escherichia_coli_UMN026_NC_011751

Escherichia_coli_O83_H1_str_NRG_857C_NC_017634

Escherichia_coli_O104_H4_str_2011C_3493_NC_018658 — *Patogénica*

2.3. Interpretar os resultados



Exercício 3: *Create ANI Matrix (ANI – Average Nucleotide Identity)*

3.1. Selecionar na barra verde do lado esquerdo a opção ***Create ANI Matrix*** no tópico *Phylogeny*.

3.2. Selecionar as estirpes:

ALL_Escherichia_coli_K_12_strain_C3026_NZ_CP014272

ALL_Escherichia_coli_O157_H7_str_Sakai_NC_002695

ALL_Escherichia_coli_O104_H4_str_2011C_3493_NC_018658

ALL_Escherichia_fergusonii_ATCC_35469_strain_ATCC_35469T_NC_011740

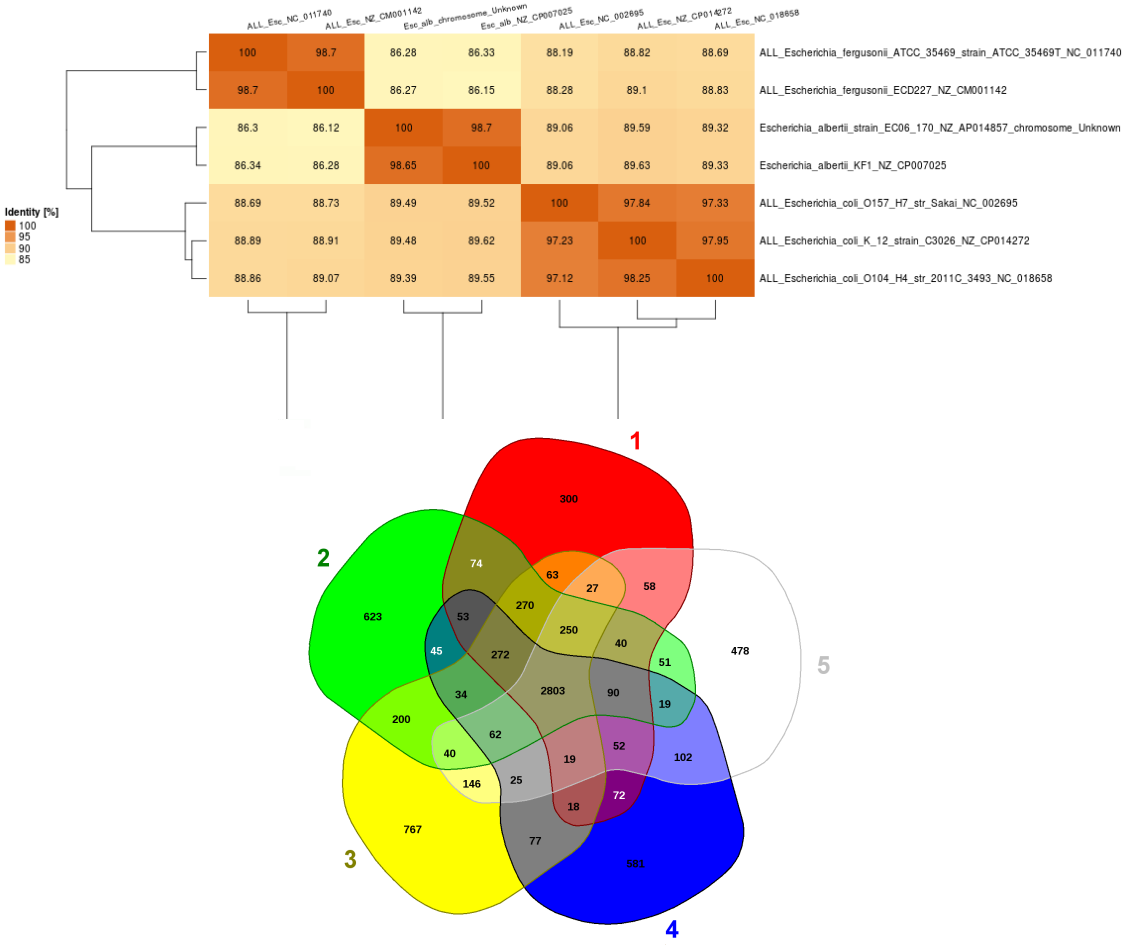
ALL_Escherichia_fergusonii_ECD227_NZ_CM001142

Escherichia_albertii_KF1_NZ_CP007025

Escherichia_albertii_strain_EC06_170_NZ_AP014857_chromosome_Unknown

3.3. Analisar os resultados.

Valor referência: Genomas que partilhem percentagens superiores a 95% são considerados pertencentes à mesma espécie (Konstantinidis & Tiedje, 2005).



1: Escherichia_coli_K_12_strain_C3026_NZ_CP014272
2: Escherichia_coli_O104_H4_str_2011C_3493_NC_018658
3: Escherichia_coli_O157_H7_str_Sakai_NC_002695
4: Escherichia_fergusonii_ATCC_35469_strain_ATCC_35469T_NC_011740
5: Escherichia_albertii_KF1_NZ_CP007025

Exercício 4: *Árvores Filogenéticas*

4.1. Selecionar na barra verde do lado esquerdo da página a opção ***Phylogentic Tree*** no tópico *Phylogeny*

4.2. Selecionar as estirpes:

ALL_Escherichia_coli_K_12_strain_C3026_NZ_CP014272

ALL_Escherichia_coli_O157_H7_str_Sakai_NC_002695

ALL_Escherichia_coli_O104_H4_str_2011C_3493_NC_018658

ALL_Escherichia_fergusonii_ATCC_35469_strain_ATCC_35469T_NC_011740

ALL_Escherichia_fergusonii_ECD227_NZ_CM001142

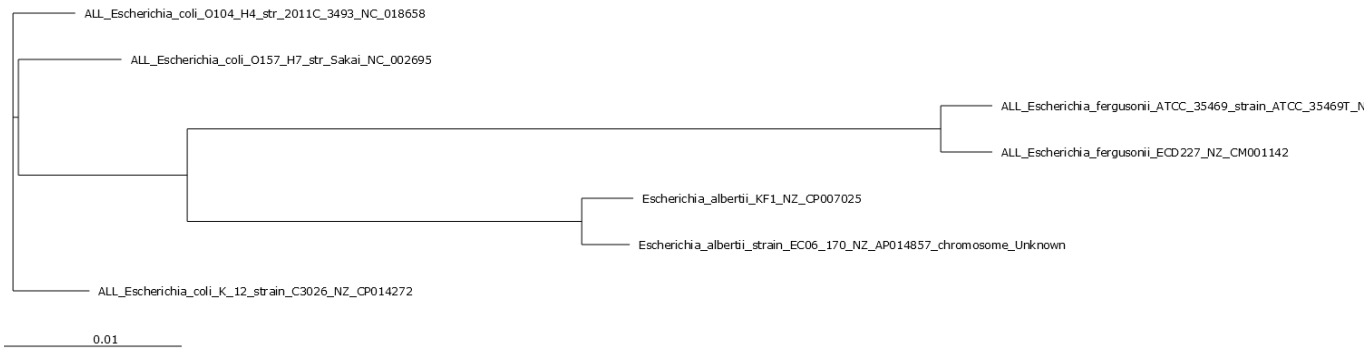
Escherichia_albertii_KF1_NZ_CP007025

Escherichia_albertii_strain_EC06_170_NZ_AP014857_chromosome_Unknown

Deve selecionar uma das estirpes anteriores como estirpe referência. Para isso clicar numa das estirpes com o botão direito do rato. A sombra sobre o nome da estirpe deve ficar verde escura e não verde clara.

4.3. Analisar os resultados.

ALL_Escherichia_coli_K_12_strain_C3026_NZ_CP014272 - estirpe de referência para esta árvore filogenética.



Appendix III: UJr Manual for Participants

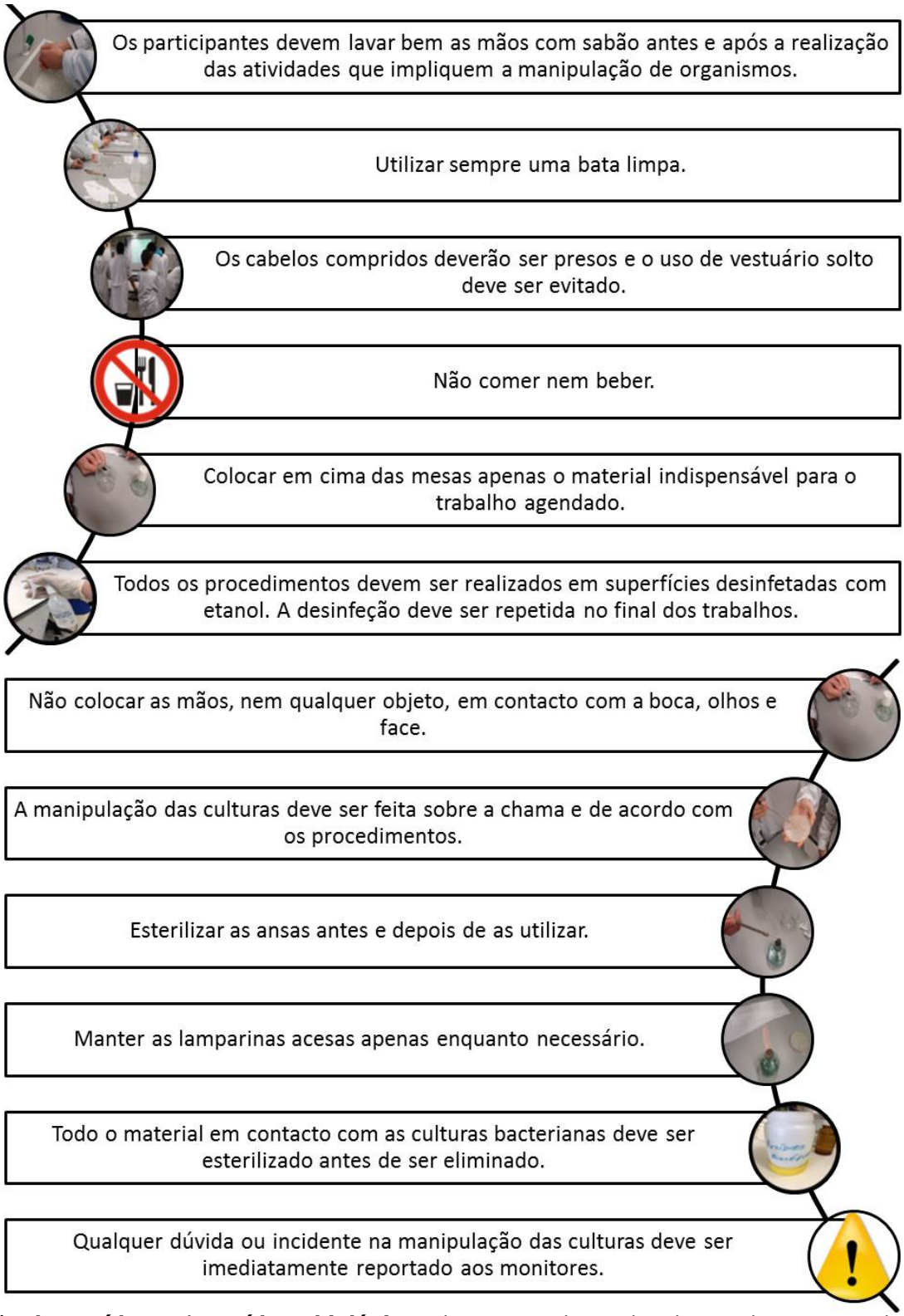
Universidade Júnior: Manual for participants

Planificação

2ª Feira	**Introdução ao trabalho laboratorial em Microbiologia** - Segurança no laboratório. - Registos de laboratório. **AP1:** Preparação, esterilização e plaqueamento de um meio de cultura.
3ª Feira	**AP2:** Pesquisa de bactérias em manipuladores e superfícies. **AP3**: Pesquisa de bactérias em alimentos.
4ªFeira	**AP4:** Avaliação das propriedades antimicrobianas do alho. **AP5:** Inoculação de meios de cultura e antibiogramas.
5ªFeira	**AP2 (*cont.*):** Observação, registo e discussão dos resultados. **AP4 (*cont.*):** Observação, registo e discussão dos resultados. **AP5 (*cont.*):** Análise e interpretação de antibiogramas. Registo e discussão dos resultados. **AP6:** Microbiologia Preditiva: Plataforma Bioinformática *ComBase* e Preservação de Alimentos.
6ªFeira	**AP5 (*cont.*):** Observação, registo e discussão dos resultados. Elaboração de poster sobre a atividade AP6. **AP7:** Uma abordagem bioinformática à evolução de resistência a antibióticos.

Segurança no Laboratório

Trabalho Laboratorial em Microbiologia



Eliminação de Resíduos: i) **Resíduos biológicos** devem ser eliminados de acordo com as indicações dos monitores, seja através de uma solução de lixívia 20%, seja através de um saco para resíduos biológicos; ii) **Resíduos não biológicos** (papel absorvente sem material biológico, fósforos, etc) devem ser colocados nos cestos do lixo e não nas tinas de material que se encontram nas mesas; iii) **Material de laboratório sujo mas não contaminado com resíduos biológicos** deve ser colocado para lavar nas tinas apropriadas.

Registos de Laboratório

Na elaboração do relatório de uma atividade experimental, deve ser contemplada a necessidade de incluir os seguintes elementos:

Título	Identificação da atividade realizada
Autor(es)	Identificação dos responsáveis pela atividade e pela elaboração do relatório.
Resumo	Breve texto descritivo dos resultados obtidos e das conclusões obtidas, destacando os aspetos de maior importância e interesse.
Objetivo	Descrição do pretendido com o trabalho executado.
Introdução	Compilação estruturada dos conhecimentos que justificam o enquadramento da atividade.
Palavras-chave	Termos relacionados com a atividade.
Protocolo experimental	Apresentação do procedimento realizado.
Resultados	Registo das observações efetuadas e dos dados recolhidos.
Discussão	Análise crítica dos resultados obtidos e seu enquadramento à luz do conhecimento actual.
Conclusão	Descrição do cumprimento ou incumprimento dos objetivos propostos para a atividade.
Comentário	Comentário global sobre a atividade experimental com recomendações para a melhoria da mesma.
Bibliografia	Listagem das fontes bibliográficas utilizadas pelo(s) autor(es).

Atividade Prática 1 (AP1): Preparação, esterilização e plaqueamento de um meio de cultura

Objetivo: Preparar, esterilizar e plaquear meios de cultura a usar nas atividades práticas.

A. Preparação de meio de cultura

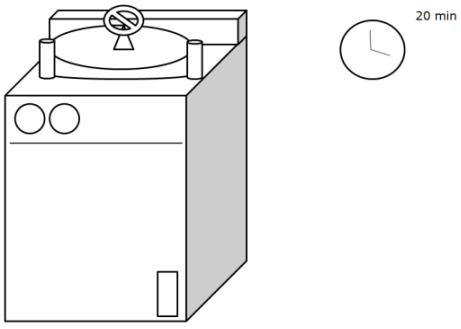
Para 1000 mL de meio Nutrient Agar *(NA)*:
1. Dissolver 23 g de meio NA em água destilada;
2. Acrescentar 3 g de agar.

Para 1000 mL de meio Muller-Hinton sólido (*MH*):
1. Dissolver 21 g de meio MH em água destilada;
2. Acrescentar 18 g de agar.

Para 1000 mL de meio Muller-Hinton líquido (MH):
1. Dissolver 21 g de meio MH em água destilada.

Para 1000 mL de meio Slanetz-Bartley (*SB*):
1. Dissolver 42 g de meio SB em água destilada.
2. Acrescentar 5 g de agar.

Para 1000 mL de meio MacConkey (*MC*):
1. Dissolver 50 g de meio MC em água destilada;
2. Acrescentar 18 g de agar.



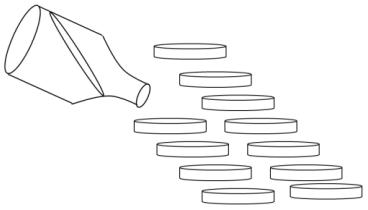
Para 1000 mL de meio Manitol e Sal (*MS*):
1. Dissolver 111 g de meio MS em água destilada;
2. Acrescentar 3 g de agar.

B. Esterilização dos meios de cultura

1. Autoclavar o meio de cultura durante a 121ºC, 1atm e durante 20 minutos.

C. Plaqueamento do meio de cultura

1. Na câmara de fluxo laminar, plaquear o meio de cultura nas placas de Petri, quando este estiver aproximadamente a 50^0 C;
2. Deixar o agar solidificar e depois identificar a placa com o nome do meio de cultura e a data da sua elaboração;
3. Inverter as placas e guardar no frigorífico.



Atividade Prática 2 (AP2): Pesquisa de bactérias em manipuladores e superfícies

Objetivo: Pesquisar a existência de bactérias em diferentes superfícies e nos manipuladores, fazendo uma inoculação de meios específicos.

A. Fornecida a Placa de Petri com meio de cultura NA (Nutrient Agar), Meio Manitol e Sal (MS), Meio Slanetz-Bartley (SB) e Meio de MacConkey (MC), proceder do seguinte modo:

1. Passar uma zaragatoa previamente humedecida com água esterilizada ou solução salina na mucosa da cavidade nasal.
2. Proceder à cultura da amostra passando a zaragatoa sobre a superfície do meio de cultura.
3. Incubar a placa na estufa à temperatura de 37° C.

B. Fornecida a Placa de Petri com meio de cultura NA (Nutrient Agar), Meio Manitol e Sal (MS), Meio Slanetz-Bartley (SB) e Meio de MacConkey (MC), proceder do seguinte modo:

1. Dividir a Placa de Petri ao meio, pelo lado exterior (plástico) com o auxílio de um marcador.

2. Passar uma zaragatoa previamente humedecida com água esterilizada ou solução salina pela superfície a analisar numa área de aproximadamente 3 cm^2.

3. Proceder à cultura da amostra passando a zaragatoa sobre metade da superfície do meio de cultura.

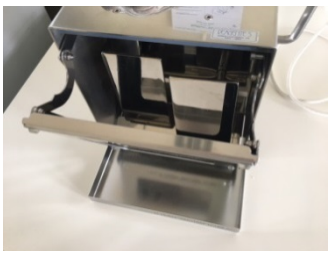
4. Repetir o passo 2 e 3 mas analisando outra superfície.

5. Incubar a placa na estufa à temperatura de 37° C.

Atividade Prática 3 (AP3): Pesquisa de bactérias em alimentos

Objetivo: Analisar a carga microbiana de um alimento usando meios de cultura seletivos de forma a estimar a diversidade e densidade microbiana do alimento. Uma contagem total baixa não assegura que um alimento esteja isento de patogénicos ou toxinas, nem uma contagem total alta significa inevitavelmente a presença de bactérias patogénicas.

A. Preparação da suspensão mãe (Diluição 10^{-1}): Músculo de Frango



1. Assepticamente pesar 10 g de músculo de frango e juntar 90 mL de solução diluente (NaCl 0,9% - soro fisiológico), num saco de homogeneização (BagFilter). Homogeneizar num BagMixer 400VW, i.e. homogeneizador tipo stomacher.

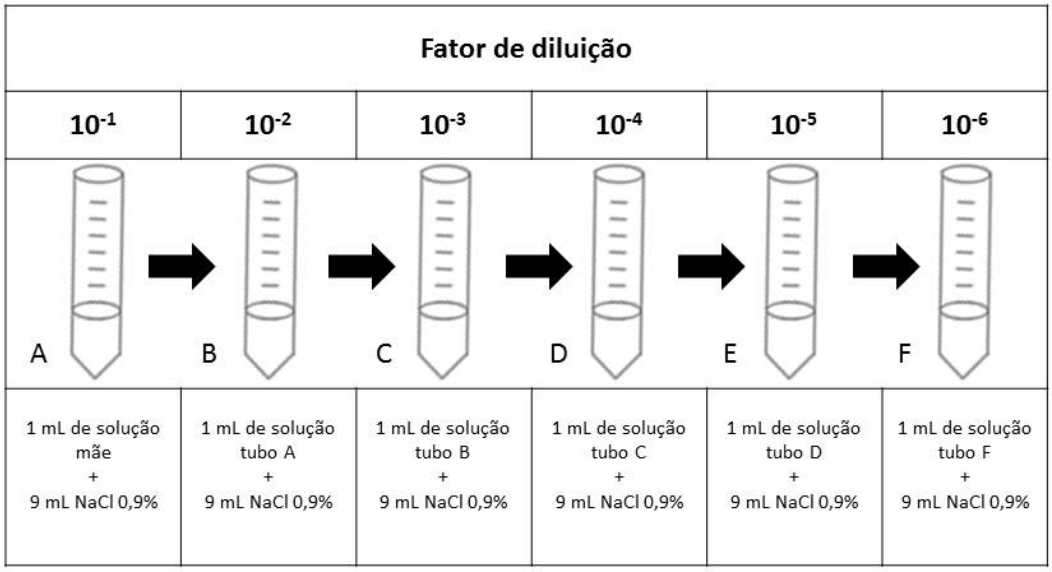
B. Preparação da suspensão mãe (Diluição 10^{-1}): Pele de Frango

1. Assepticamente pesar 10 g de pele de frango e juntar 90 mL de solução diluente (NaCl 0,9% - soro fisiológico), como solução diluente, num saco de homogeneização (BagFilter). Homogeneizar num BagMixer 400VW, i.e. homogeneizador tipo stomacher.

C. Preparação de diluições decimais

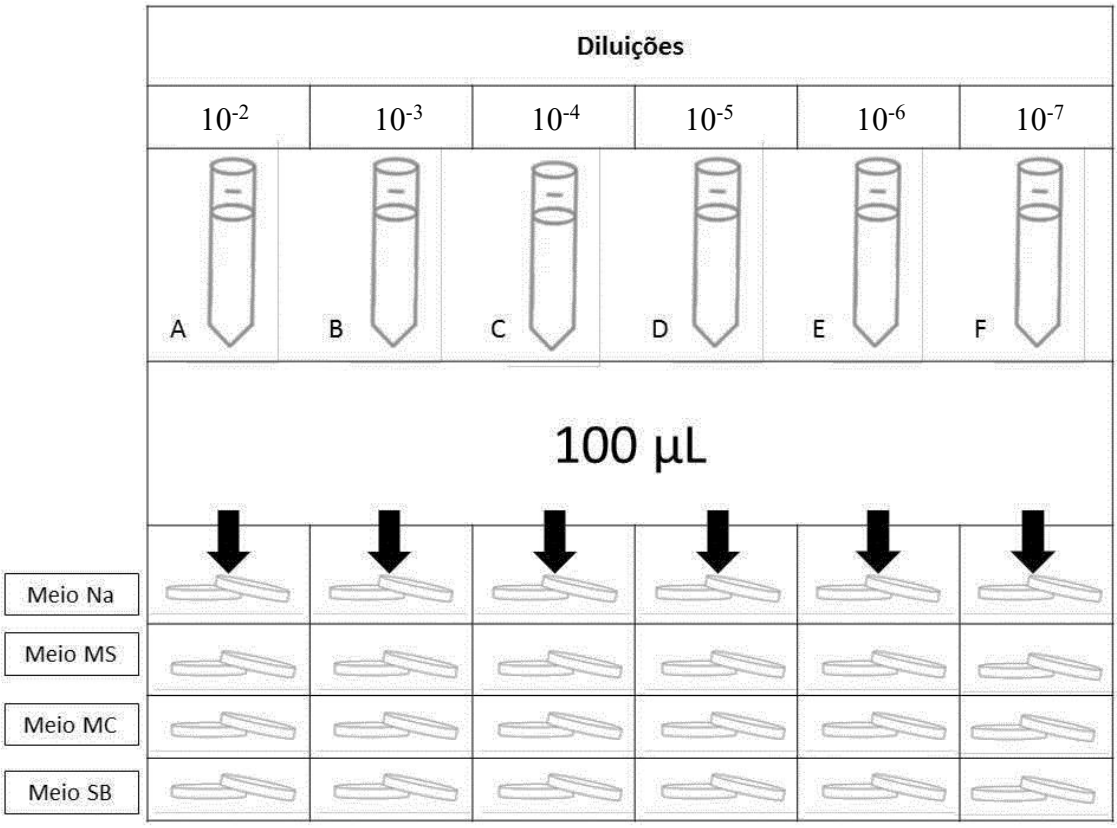
1. Preparar, a partir de cada uma das suspensões mãe (músculo e pele de frango), diluições decimais em soro fisiológico estéril (diluir 1 mL da suspensão mãe em 9 mL de NaCl 0,9%).

2. Preparar diluições de acordo com o esquema indicado na página seguinte:



D. Fornecida a Placa de Petri com o meio de cultura NA, Meio Manitol e Sal, e Meio de MacConkey, proceder do seguinte modo:

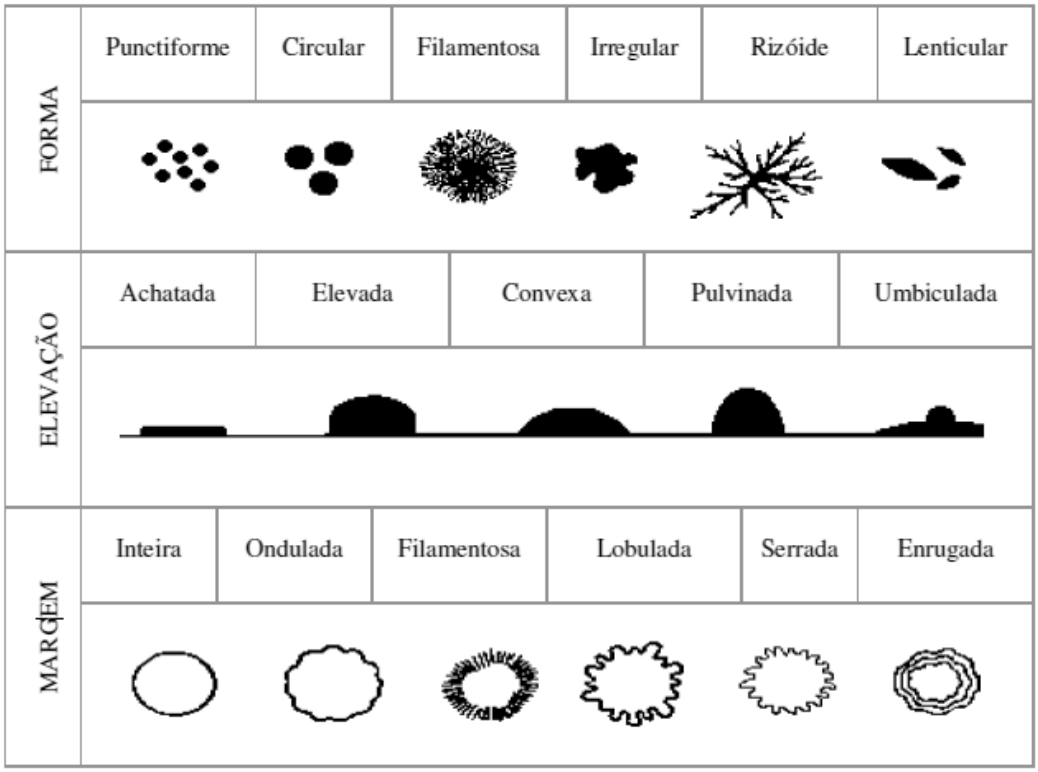
1. Pipetar 100µL de cada diluição e proceder à cultura da amostra utilizando um espalhador para distribuir a amostra pela superfície do meio de cultura de acordo com o esquema da tabela.



2. Colocar a placa invertida na estufa a incubar à temperatura de 37° C.

Após incubação, registar os resultados obtidos, tendo em atenção os seguintes aspetos:

- Comprovar o crescimento de microrganismos através do aparecimento de colónias à superfície do meio sólido. Estudar a diversidade morfológica microbiana (bactérias e fungos) por observação do fenótipo das colónias.



- Registar o número de UFC (unidades formadoras de colónias) em cada caixa e quando possível calcular o número de UFC por cm^2 da superfície estudada.

- Quando pertinente comparar o resultado obtido com a legislação em vigor e tirar as devidas conclusões.

Atividade Prática 4 (AP4): Avaliação das propriedades antimicrobianas do alho

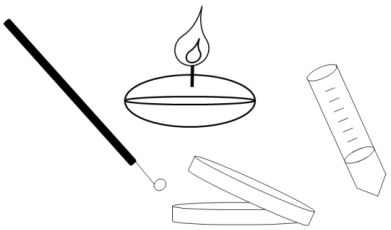
Objetivo

Avaliar o efeito antibacteriano de extrato aquoso de alho sobre *Bacillus* sp, através do método de difusão em meio sólido.

Procedimento

Preparação do inóculo bacteriano

- Preparar suspensões de *Bacillus* sp, transferindo uma pequena quantidade de células de uma cultura em placa de meio NA para um tubo com 5 mL de água destilada esterilizada, e agitando para homogeneizar o conteúdo.

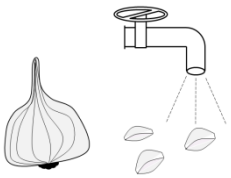


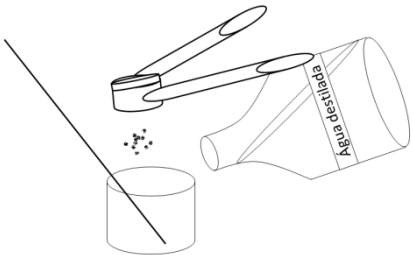
Preparação do extrato aquoso de alho

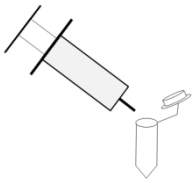
- Descascar e lavar em água corrente três ou quatro dentes de alho de tamanho médio.

- Esmagar 3 a 5 g de dentes de alho num almofariz esterilizado e adicionar a mesma proporção de água destilada esterilizada, 3 a 5 mL (1:1).

- Filtrar o homogeneizado para tubos esterilizados através de uma seringa de 10 mL contendo gaze.





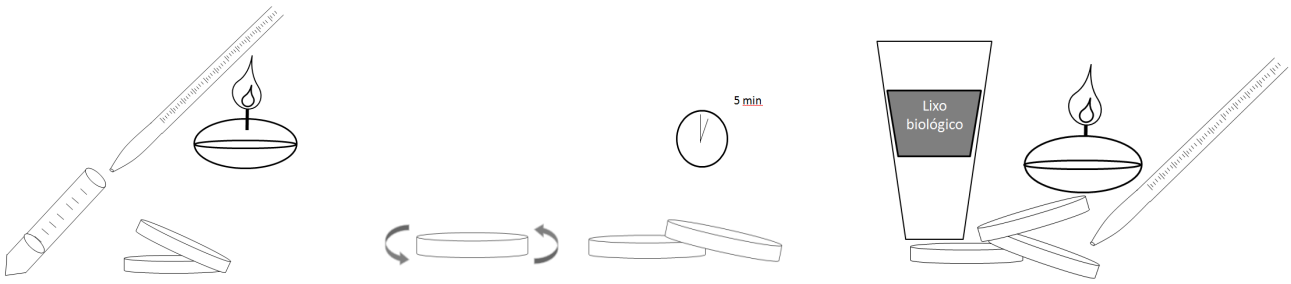


Preparação das placas

- Transferir 1 mL de suspensão bacteriana para uma placa de meio.

- Rodar cuidadosamente a placa fechada para distribuir a suspensão por toda a superfície do meio de cultura.

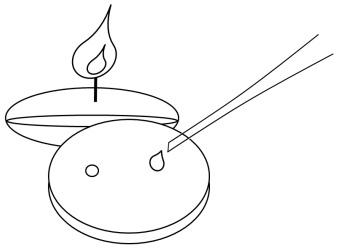
- Deixar a placa inclinada durante cinco minutos e remover o excesso de suspensão com uma micropipeta.



- Deixar as placas repousar sem inverter durante cerca de cinco minutos. Legendar as placas.

4. Bioensaio

- Com uma micropipeta aplicar 10 µL (= 0.01 mL) de água destilada esterilizada num dos lados de cada placa (controlo). Aplicar 10 µL de extrato aquoso de alho no outro lado.



- Deixar as placas repousar sem inverter até a água e o extrato se difundirem no meio de cultura.

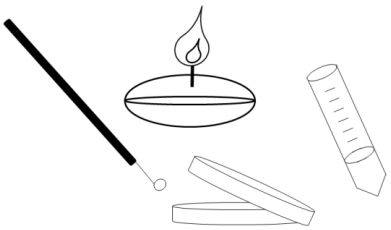
- Incubar as placas em posição invertida durante 24 horas a 37^0 C.

- Observar, registar e interpretar os resultados.

<u>Atividade Prática 5</u> (AP5): Inoculação de meios de cultura e antibiogramas

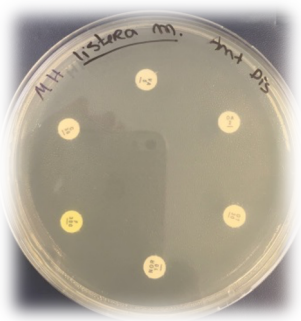
Objetivo: Identificar a resistência uma bactéria específica a diferentes antibióticos.

1. Preparar suspensões da bactéria a estudar (isolado ou cultura a fornecer), transferindo uma pequena quantidade de células de uma cultura em placa para um tubo com 5 mL de meio líquido Muller–Hinton (MH) e agitar para resuspender as células.



2. Humedecer uma zaragatoa na suspensão bacteriana preparada;
3. Proceder à sementeira da amostra passando a zaragatoa sobre a superfície do meio de cultura (Mueller-Hinton).
4. Colocar os discos de antibiótico com o dispensador.





5. Colocar a placa na estufa a incubar à temperatura de 37° C.

Análise e interpretação de antibiogramas

Objetivo: Analisar o efeito dos discos de antibiótico na inibição do crescimento da bactéria estudada.

1. Observar as placas e identificar a ocorrência de zonas de inibição do crescimento bacteriano em torno dos pontos de aplicação do(s) antibiótico(s).
2. Medir o diâmetro do(s) halo(s) observado(s) e interpretar os resultados, de acordo com a informação disponível na tabela seguinte.

Antibiótico	Modo de ação	Conteúdo do Disco (µg)	Interpretação do antibiograma (diâmetro do halo, mm)	
			Suscetível	Resistente
Ampicilina	*Bactericida*	**AM10**		
Enterobacteriaceae	Interfere com a síntese da parede celular.	10	≥ 17	≤ 13
Estafilococos		10	≥ 29	≤ 28
Enterococos		10	≥ 17	≤ 16
Estreptococos (sem *S. pneumoniae*)		10	≥ 26	≤ 18
Trimetoprim-Sulfametoxazol	*Bacteriostáticos*	**SXT25**		
Organismos (sem *S. pneumonia)*	Inibem a síntese de ácido fólico.	1.25/23.75	≥ 16	≤ 10
Streptococcus pneumoniae		1.25/23.75	≥ 19	≤ 15
Tetraciclina	*Bacteriostático (bactericida em altas concentrações)*	**TE10**		
Organismos (sem Estreptococos)	Interfere com a síntese proteica.	30	≥ 19	≤ 14
Estreptococos		30	≥ 23	≤ 18
Vancomicina	*Bactericida*	**VA10**		
Enterococos	Interfere com a síntese da parede celular.	30	≥ 17	≤ 14
Estreptococos		30	≥ 17	____
Outros organismos gram-positivos		30	≥ 12	≤ 9
Estreptomicina [a]	*Bacteriostático*	**S10**		
Enterobacteriaceae	Interfere com a síntese proteica.	10	≥ 15	≤ 11
Eritromicina	*Bacteriostático*	**E10**		
Organismos (sem Estreptococos)	Interfere com a síntese proteica.	15	≥ 23	≤ 13
Estreptococos		15	≥ 21	≤ 15

(Informação obtida a partir das tabelas de Breakpoint para interpretação dos halos de inibição do EUCAST (2013); e do CLSI[a] (documento M100-S17, 2007).

Atividade Prática 6 (AP6): *Microbiologia Preditiva*

Objetivo: Compreender o efeito de fatores extrínsecos (como a temperatura e concentração de oxigénio) e de fatores intrínsecos (como o pH e osmolaridade), no crescimento de diferentes bactérias em alimentos.

Compreender a importância destes dados para a optimização de métodos de preservação de alimentos.

A. Modelos de crescimento

1. Aceder à plataforma digital *ComBase*: http://www.combase.cc/index.php/en/

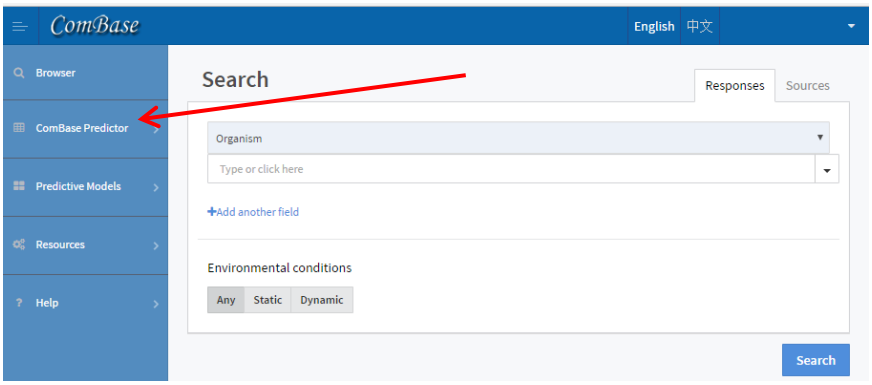
2. Fazer o *Login/Register*

3. Clicar em Sign Up

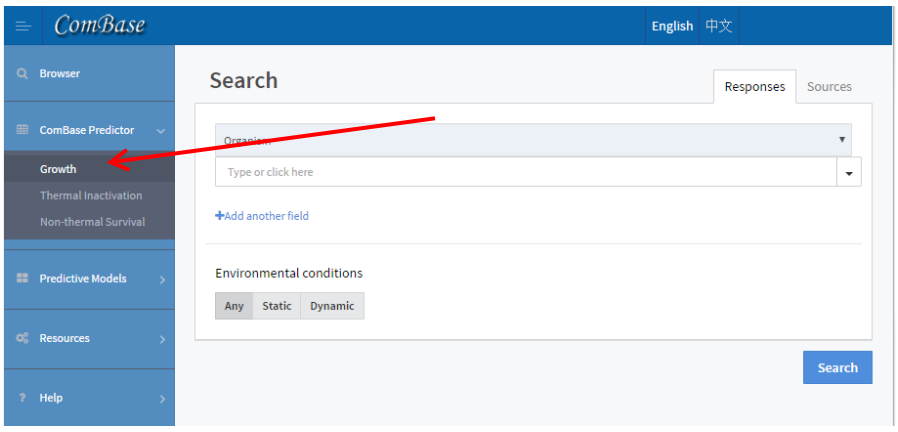
4. Efetuar o registo

5. Fazer o *Log in* na plataforma com os dados de registo (recebidos no email):

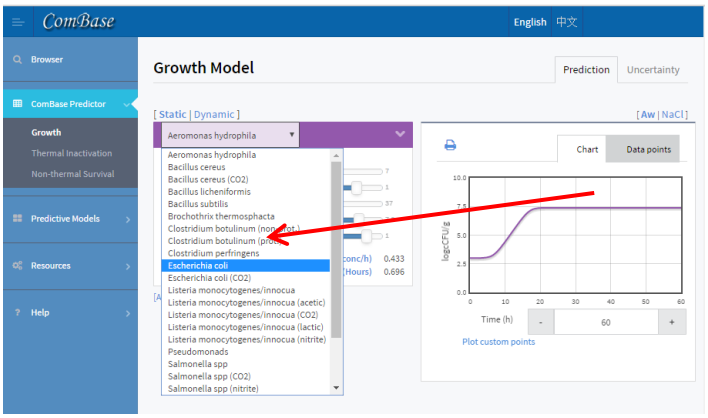
6. Aceder ao menu *ComBase Predictor*



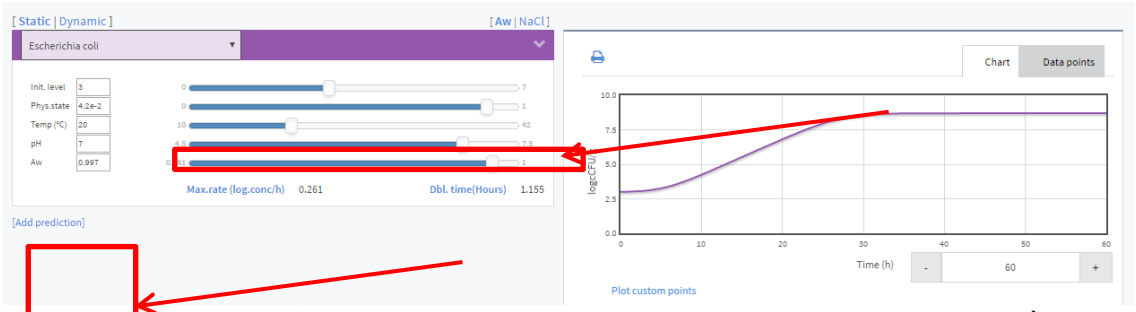
7. Selecionar o modelo de crescimento (*Growth*)



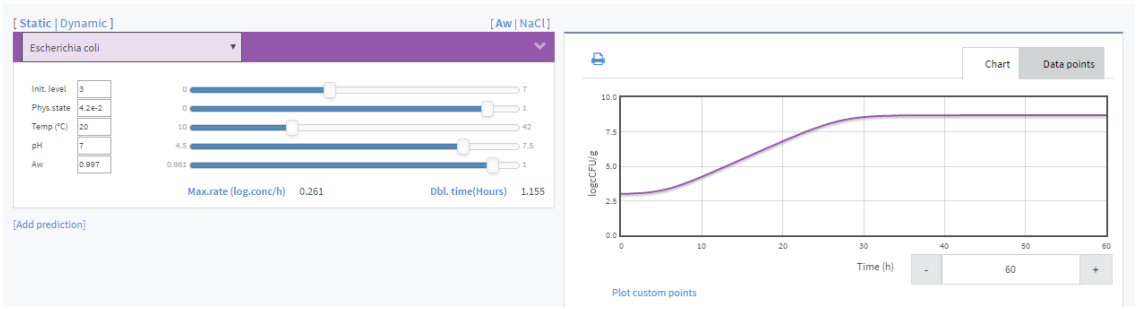
8. No organismo a estudar escolher *Escherichia coli*



9. Identificar as diferentes fases de crescimento bacteriano e analisar os parâmetros: *Max. Rate* (Taxa de crescimento) e *Dbl.time* (Tempo de duplicação)



10. Alterar as condições em que a *Escherichia coli* se encontra e responder às questões **a** e **b**, justificando com dados recolhidos na plataforma, nomeadamente o *Max. Rate* e o *Dbl.time*:



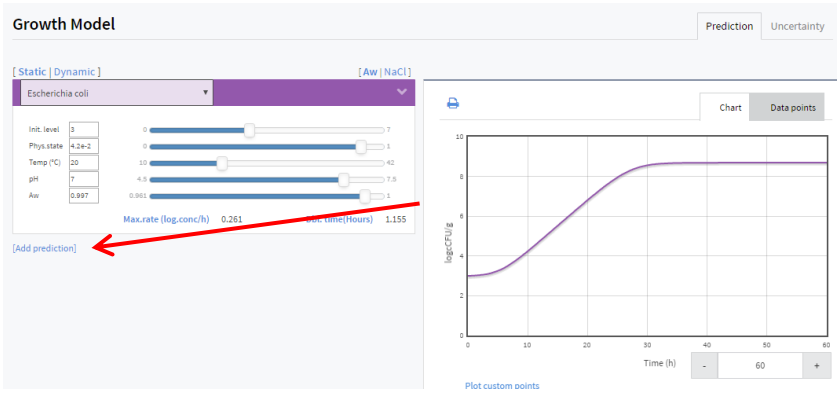
11. Responde às seguintes perguntas:

a) Como varia o crescimento da bactéria *Escherichia coli* a temperaturas de refrigeração (a temperatura nos frigoríficos domésticos é de ±6ºC), e à temperatura ambiente (±20˚C). Em que condições o crescimento da bactéria é mais lento?

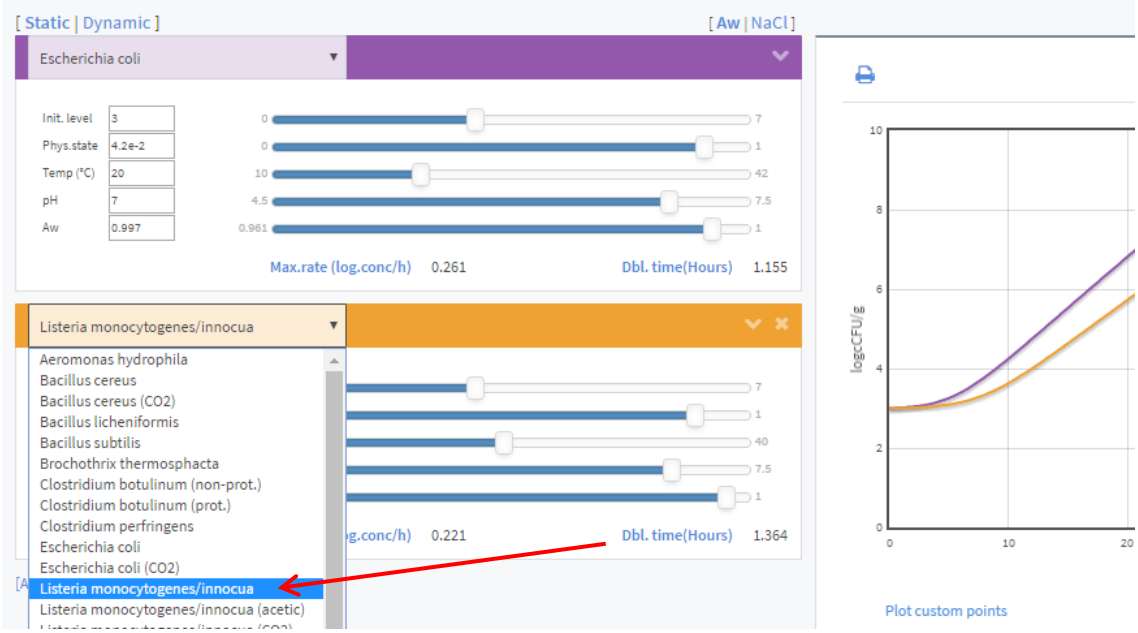
b) Sabendo que temperaturas de refrigeração (±6ºC) implicam um enorme consumo de energia, tornando o processo de preservação de alimentos muito dispendioso, que outros fatores podem ser alterados (exemplo o pH e a concentração de cloreto de sódio - NaCl) de modo a que, mantendo os alimentos à temperatura ambiente (±20˚C), as culturas de *Escherichia coli* mantenham um crescimento igual ou inferior ao obtido à temperatura de refrigeração escolhida.

B. Comparar o crescimento de duas espécies bacterianas

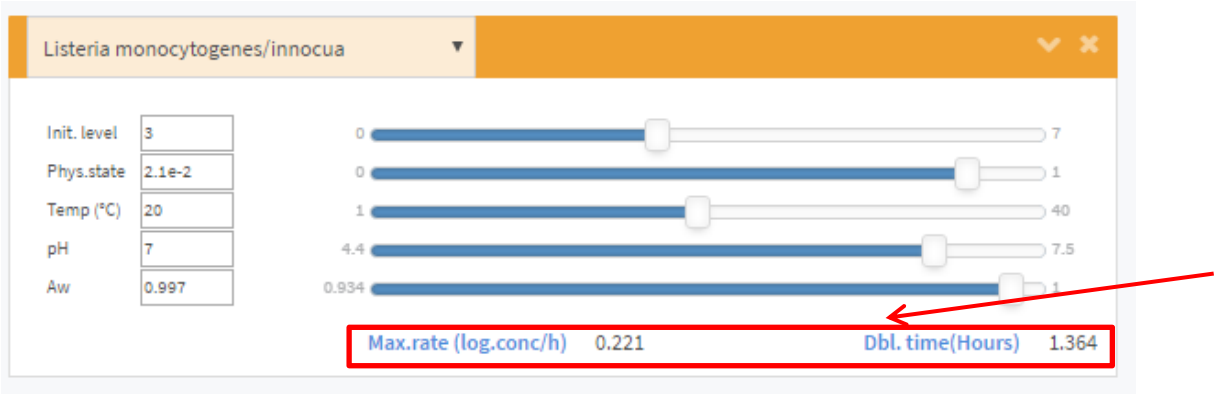
1. Selecione a opção *Add prediction*



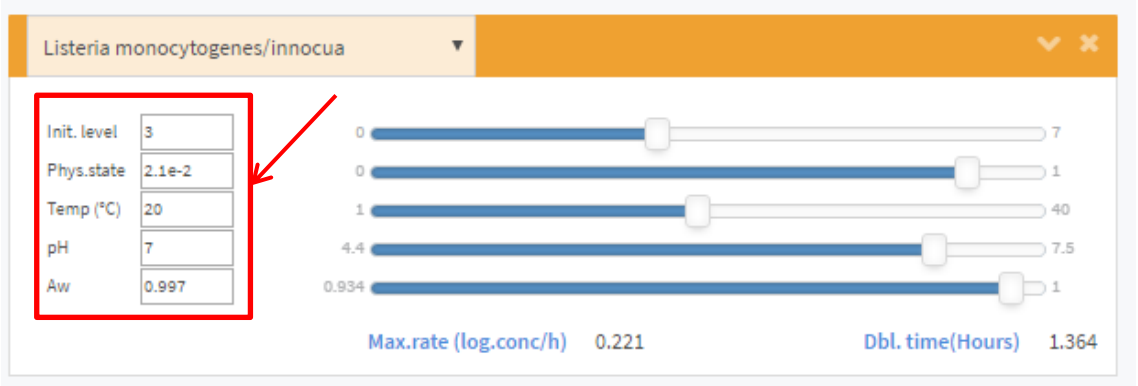
2. Escolha a espécie *Listeria monocytogenes*



3. Identificar as diferentes fases de crescimento bacteriano e analisar os parâmetros: Max. Rate e Dbl.time



4. Alterar as condições em que *Listeria monocytogenes* se encontra e responder às questões **a** e **b** (página seguinte), justificando com dados recolhidos na plataforma, nomeadamente o Max. Rate e o Dbl.time:



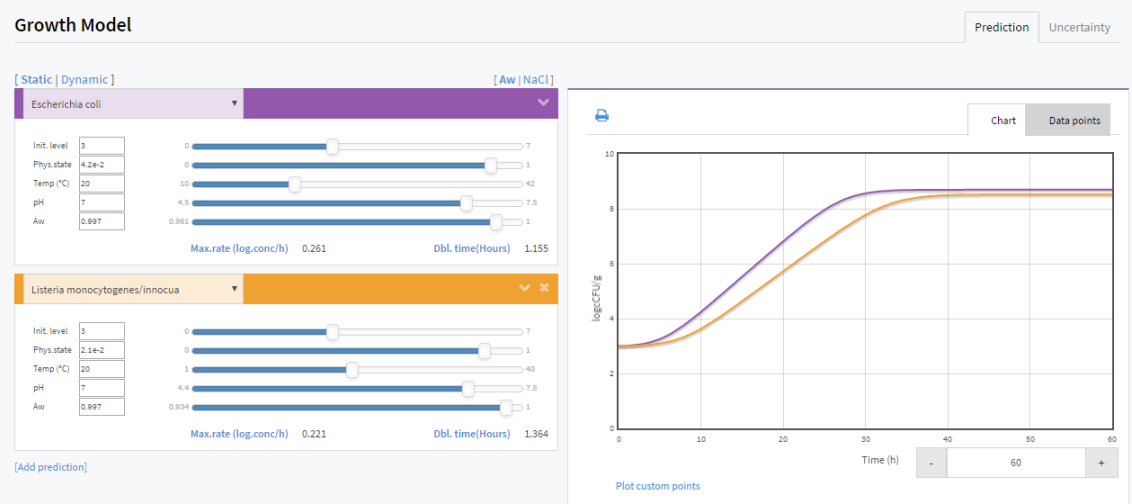
5. Questões:

a) Como varia o crescimento da bactéria *Listeria monocytogenes* a temperaturas de refrigeração (a temperatura nos frigoríficos domésticos é de ±6ºC), e à temperatura ambiente (±20°C). Em que condições o crescimento da bactéria é mais lento?

b) Sabendo que temperaturas de refrigeração (±6ºC) implicam um enorme consumo de energia, tornando o processo de preservação de alimentos muito dispendioso, que outros fatores podem ser alterados (exemplo o pH e a concentração de cloreto de sódio - NaCl) de modo a que, mantendo os alimentos à temperatura ambiente (±20°C), as culturas de *Listeria monocytogenes* mantenham um crescimento igual ou inferior ao obtido à temperatura de refrigeração escolhida.

Modelos de inibição do crescimento bacteriano em E. coli e L. monocytogenes

Podendo comparar o crescimento das duas espécies, estude as condições relativas às variáveis: temperatura, pH e Concentração em NaCl, com o objetivo de inibir o crescimento das duas espécies por mais tempo.



**Nota explicativa:

Phys. state ("*initial physiological state*"/estado fisiológico inicial): o valor deste parâmetro pode variar entre 0 e 1 e expressa a adequação metabólica das bactérias ao ambiente. Se o seu valor for 0, então o crescimento não ocorrerá (fase *lag* infinita); Se o valor for 1, então o crescimento começará imediatamente (sem fase *lag*).

Atividade Prática 7 (AP7): Abordagem bioinformática à evolução de resistência a antibióticos

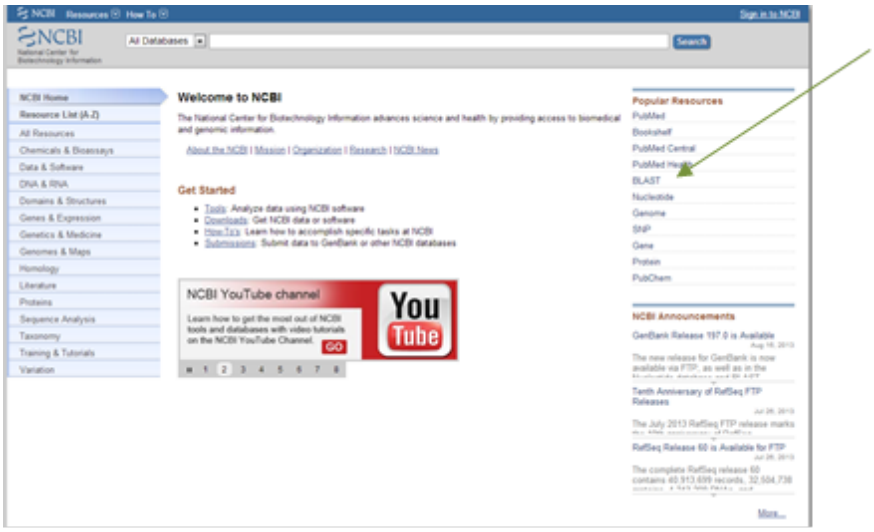
Objetivo: Abordar a resistência a antibióticos através de uma análise bioinformática dos genes de diferentes bactérias.

A seguinte sequência de aminoácidos corresponde a uma proteína responsável pela resistência ao antibiótico tetraciclina no organismo *Ureaplasma urealyticum*:

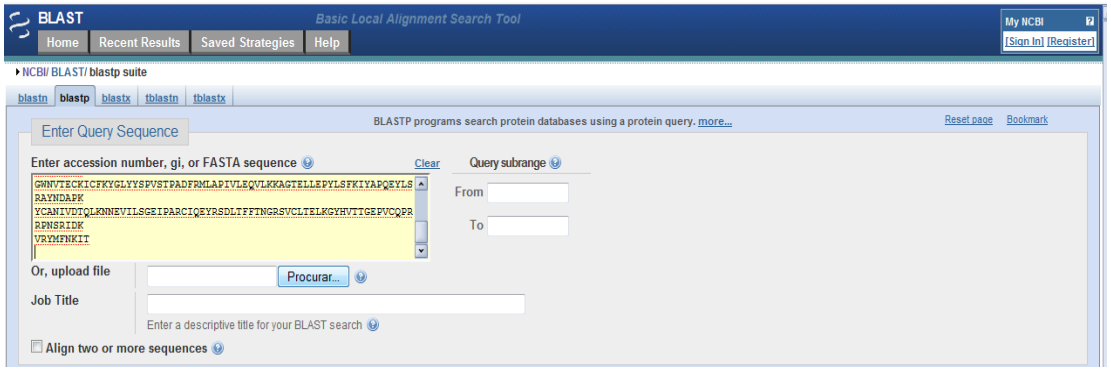
>**gi|475984**|gb|AAA73978.1| tetracycline resistance determinant [Ureaplasma urealyticum]
MKIINIGVLAHVDAGKTTLTESLLYNSGAITELGSVDKGTTRTDNTLLERQRGITIQTGITSFQWEN
TKV
NIIDTPGHMDFLAEVYRSLSVLDGAILLISAKDGVQAQTRILFHALRKMGIPTIFFINKIDQNGIDLST
V
YQDIKEKLSAEIVIKQKVELYPNMCVTNFTESEQWDTVIEGNDDLLEKYMSGKSLEALELEQEESI
RFHN
CSLFPVYHGSAKNNIGIDNLIEVITNKFYSSTHRGPSELCGNVFKIEYTKKRQRLAYIRLYSGVLHL
RDS
VRVSEKEKIKVTEMYTSINGELCKIDRAYSGEIVILQNEFLKLNSVLGDTKLLPQRKRIENPHPLLQI
TV
EPSKPEQREMLLDALLEISDSDPLLRYYVDSTTHEIILSFLGKVQMEVISALLQEKYHVEIELKEPTV
IY
MERPLKNAEYTIHIEVPPNPFWASIGLSVSPLPLGSGMQYESSVSLGYLNQSFQNAVMEGIRYGC
EQGLY
GWNVTECKICFKYGLYYSPVSTPADFRMLAPIVLEQVLKKAGTELLEPYLSFKIYAPQEYLSRAYN
DAPK
YCANIVDTQLKNNEVILSGEIPARCIQEYRSDLTFFTNGRSVCLTELKGYHVTTGEPVCQPRRPNS
RIDK
VRYMFNKIT

1. Aceder à página do National Center for Biotechnology Information (NCBI).

2. Aceder à ferramenta Blast [esta ferramenta permite, através de um algoritmo matemático, localizar as sequências mais semelhantes à sequência em estudo presentes na base de dados.



3. Seleccionar "protein blast", e colocar o *accession number* (**gi|475984**) da sequência em estudo no campo "query sequence".



4. Iniciar a busca clicando em "Blast" (no final da página).

5. Verificar a quais organismos pertencem as 10 proteínas mais semelhantes à proteína em estudo.

6. Voltar ao formulário do Blast e escrever "Tenericutes" no campo "Entrez query" (dessa forma, a pesquisa ficará limitadas ao filo Tenericutes, ao qual pertence *Ureaplasma urealyticum*).

7. Fazer nova busca clicando em "Blast".

8. Verificar a quais organismos, dentro dos Tenericutes, pertencem as 10 proteínas mais semelhantes à proteína em estudo.

Comparar o primeiro resultado com o segundo. O que se pode concluir?

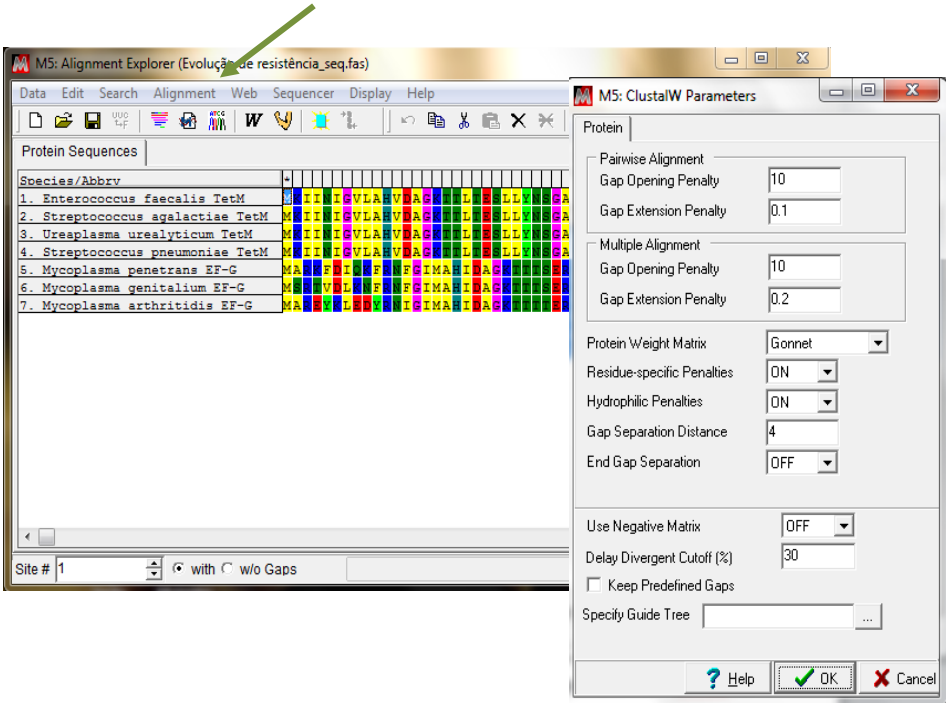
1. Evolução da resistência a antibióticos

- No computador, abrir o ficheiro seq.fas utilizando o programa Mega (fazer duplo click sobre o ficheiro).

Este ficheiro contém a proteína original, TetM de *U. urealitycum*, que faz parte do mecanismo de resistência ao antibiótico tetraciclina, bem como as proteínas homólogas (isto é, com a mesma origem) elongation-factor G. Estas proteínas EF-G são responsáveis pela translocação do mRNA e dos tRNAs ao longo dos ribossomas durante a síntese proteica.

A tetraciclina exerce a sua acção antibiótica impedindo a ligação entre os tRNAs e os ribossomas: dessa forma, a produção proteica cessa e o organismo acaba por morrer. As proteínas TetM funcionam como protectoras dos ribossomas, possibilitando a síntese proteica mesmo na presença de tetraciclina.

1. Alinhar as sequências seleccionando todas as sequências no menu **Alignment,** seleccionar a primeira opção *Align by ClustalW/* na janela de opções, manter as opções pré-definidas e clicar em **OK**. Ver Figura demonstrativa na página seguinte.

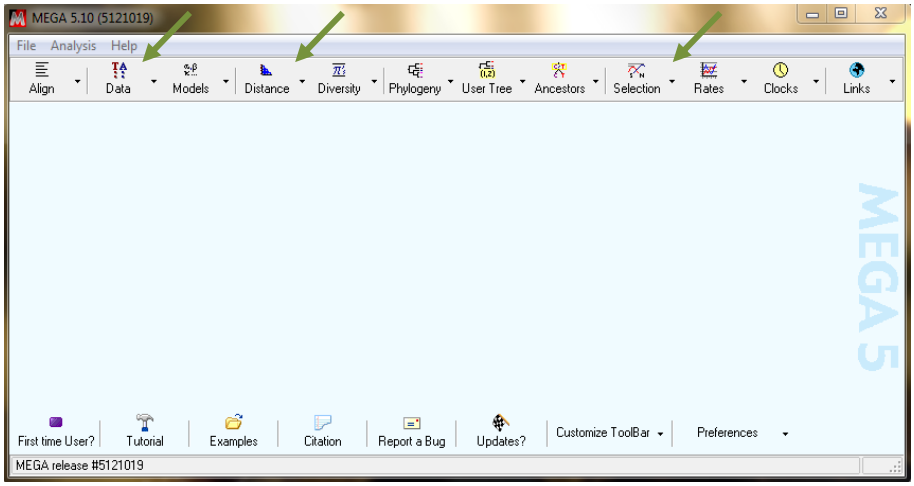


2. Observar o alinhamento. É possível verificar que há zonas completamente alinhadas (conservadas) entre todas as sequências, zonas conservadas entre algumas sequências, e zonas que são completamente diferentes em todas as sequências.

<u>O que é que a conservação de cada zona específica da proteína diz sobre a funcionalidade dessas regiões?</u>

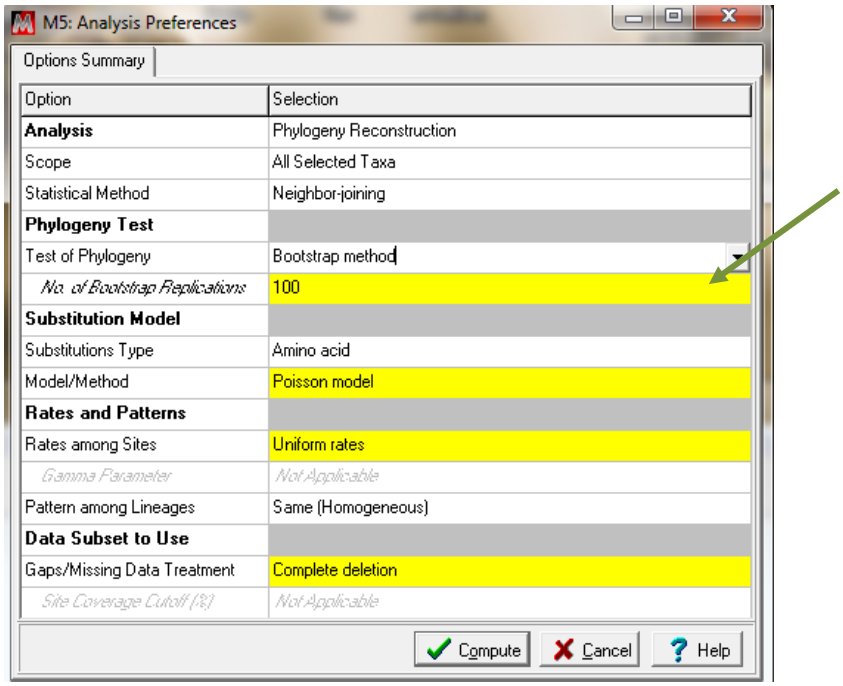
1. Exportar o alinhamento para o formato .mega, seleccionando a opção *Export Alignment>MEGA format* no menu **Data**. Ver Figura em baixo.

2. Fechar a janela do alinhamento e abrir o ficheiro exportado para .mega. Verificar o aparecimento de um conjunto de novas opções na janela inicial do Mega.



3. Inferir as relações evolutivas construindo uma árvore filogenética pelo método Neighbor-Joining a partir do alinhamento das sequências (no menu **Phylogeny**, seleccionar a opção *Construct Phylogeny>Neighbor-Joining (NJ)…*).

Na janela de opções que surge, certifica-te que na opção Phylogeny Test, está definido um Bootstrap de 100 réplicas.



4. Observar a árvore construída. Com base nos resultados obtidos no exercício nº 1 e sabendo que:

i) *Enterococcus faecalis*, *Streptococcus agalactiae* e *Streptococcus pneumoniae* pertencem ao filo **Firmicutes**;

ii) *Mycoplasma penetrans*, *Mycoplasma genitalium*, *Mycoplasma arthritidis* e *Ureaplasma urealyticum* pertencem ao filo **Tenericutes**;

Que conclusões se podem retirar sobre a evolução da proteína TetM na espécie *Ureaplasma urealyticum*?

- Embora *Ureaplasma urealyticum* faça parte da flora normal do tracto urogenital dos humanos, pode, sob determinadas condições (por exemplo, imunossupressão medicamentosa ou outra) causar infecções urinárias.

Quais são as consequências da aquisição por parte destes microrganismos do gene *TetM*, nomeadamente no que diz respeito ao tratamento destas infecções?

2. Para pensar…

- Comentar a seguinte frase:

Os mecanismos de resistência aos antibióticos são mais variados e estão mais disseminados nos microrganismos do que a produção de antibióticos, que está restrita a um pequeno grupo de microrganismos.

Tendo em consideração que a produção de antibióticos para os microrganismos produtores, representa:

i) uma vantagem evolutiva (competição com outras populações bacterianas pelo mesmo nicho ecológico);

ii) um custo metabólico elevado (em termos de matéria prima e energia).

Os meus apontamentos …

Anexo: Modelo de Poster

Microbiologia Preditiva: Plataforma Bioinformática *ComBase* e Preservação de Alimentos

Autores

Resumo/Abstract

Neste estudo, através da plataforma bioinformática *ComBase*, foi explorada a importância de modelos preditivos para a indústria alimentar, sistematizados através de aplicações computacionais.

Palavras-Chave

Bioinformática; Indústria Alimentar; Microbiologia Preditiva; Preservação de Alimentos

Comentários

Comentários/ Reflexões sobre a atividade AP6: pontos positivos; pontos negativos; sugestões de melhoria …

Introdução

A plataforma bioinformática ComBase permite compreender o efeito de fatores extrínsecos (como a temperatura e concentração de oxigénio) e de fatores intrínsecos (como o pH e osmolaridade), no crescimento de diferentes bactérias em alimentos.
Através dos dados obtidos nas simulações realizadas na plataforma é possível inferir a importância de cada um dos fatores analisados para a optimização de métodos de preservação de alimentos.

Materiais e Métodos

Ver Guia Prático da Universidade Júnior 2018: Atividade Prática 6 (**AP6**)

Resultados/Discussão

Indicação dos dados obtidos; Análise dos Resultados (explorar os diferentes modelos preditivos usados) ; Registo Gráfico…

Conclusão

Reflexão sobre a utilidades das plataformas bioinformáticas numa perspectiva da sua aplicação na indústria alimentar

Bibliografia

✓ Tavares F.; Martins A,; Dias C.; Fonseca, MJ (2018). *Bactérias, Antibióticos e Resistência: Vamos descobrir o que os liga?* – Guia de Atividade. Universidade Júnior UP.
✓ *ComBase* website: http://www.combase.cc/index.php/en/ (acedido em 15/06/2018).